

VU Research Portal

Competition, dynamic pricing and advice in frictional markets

Thiel, J.H.

2020

document version

Publisher's PDF, also known as Version of record

[Link to publication in VU Research Portal](#)

citation for published version (APA)

Thiel, J. H. (2020). *Competition, dynamic pricing and advice in frictional markets: Theory and evidence from the Dutch market for mortgages*. [PhD-Thesis - Research and graduation internal, Vrije Universiteit Amsterdam].

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:

vuresearchportal.ub@vu.nl

Competition, dynamic pricing and advice in frictional markets

Theory and evidence from the Dutch market for mortgages

Jurre H. Thiel

ISBN 978 90 3610 582 8

Cover design: Crasborn Graphic Designers bno, Valkenburg a.d. Geul

This book is no. 756 of the Tinbergen Institute Research Series, established through cooperation between Rozenberg Publishers and the Tinbergen Institute. A list of books which already appeared in the series can be found in the back.

VRIJE UNIVERSITEIT

**COMPETITION, DYNAMIC PRICING AND ADVICE IN FRICTIONAL
MARKETS**

Theory and evidence from the Dutch market for mortgages

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor of Philosophy aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. V. Subramaniam,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de School of Business and Economics
op dinsdag 14 januari 2020 om 13.45 uur
in de aula van de universiteit,
De Boelelaan 1105

door

Jurre Hendrik Thiel

geboren te Amsterdam

promotor: prof.dr. J.L. Moraga-González
copromotor: dr. M. Mastrogiacomo

Views expressed are those of the author and do not necessarily reflect official positions of De Nederlandsche Bank. In this dissertation use is made of data of the DNB Household Survey.

Contents

Contents	vii
Acknowledgments	xi
List of Figures	xiv
List of Tables	xv
Introduction	xvii
1 Price discrimination, switching costs and welfare: Evidence from the Dutch mortgage market	1
1.1 Introduction	1
1.2 Related literature	6
1.3 The Dutch mortgage market	7
1.3.1 Ban on history-based price discrimination	8
1.4 Data	9
1.4.1 Data sources	9
1.4.2 Inferring switching	10
1.4.3 Sample selection	11
1.4.4 Descriptive statistics	12
1.4.5 Evidence of history-based price discrimination	13
1.5 A model of mortgage demand and supply	16
1.5.1 Empirical strategy	16
1.5.2 Demand	17
1.5.3 Supply	20

CONTENTS

1.5.4	Sparse Markov Perfect Equilibrium	22
1.6	Identification and estimation	28
1.6.1	Sparse states and policy functions	28
1.6.2	Demand	31
1.6.3	Supply: marginal cost estimation	32
1.7	Results	35
1.7.1	Model primitives	35
1.7.2	The effect of history-based price discrimination on consumer surplus, profits and welfare	39
1.8	Discussion	43
1.8.1	Other changes in the Dutch mortgage market	43
1.8.2	Predicting pre-ban interest rates	44
1.8.3	The empirical content of SMPE	46
1.9	Conclusion	48
1.A	Testing the implications of SMPE	49
1.B	Estimation details	51
1.B.1	Matching households over time	51
1.B.2	Supply side first order conditions	52
1.B.3	Calculating a bank's best response	53
1.B.4	Calculating the cost of inattention	53
2	Competition and dynamic pricing in markets with consumer switching costs	55
2.1	Introduction	55
2.2	The effects of entry in a simple model of switching costs	59
2.3	The Dutch mortgage market	63
2.4	Data	66
2.5	The effects of entry on interest rates	68
2.5.1	Main empirical specification	68
2.5.2	Identification	70
2.5.3	Main results	75
2.5.4	Robustness	76
2.6	Conclusion	77
2.A	Additional tables and figures	78
3	Fee-based versus commission-based advice	81
3.1	Introduction	81
3.2	A model of advice	86

3.3	Short-run equilibria	89
3.3.1	Short-run equilibrium with kickbacks	89
3.3.2	Short-run equilibrium without kickbacks	94
3.4	Capacity constraints and entry	97
3.4.1	Long-run equilibrium with kickbacks	99
3.4.2	Long-run equilibrium without kickbacks	100
3.4.3	Consumer surplus with and without kickbacks	103
3.5	Extensions	105
3.5.1	Behavioral consumers	105
3.5.2	Advisers can recommend to make no purchase	107
3.5.3	Consumers do not require an adviser to make a purchase	110
3.5.4	Discussion of assumptions	112
3.6	Conclusion	113
3.A	Proofs	114
3.B	Alternative assumptions	129
3.B.1	The model without capacity constraints	129
3.B.2	Long-run equilibria without advisers' convenience benefits	131
	Summary	133
	Bibliography	135

Acknowledgments

So, this is it. More than 150 pages and 22000 words, thirteen tables, twenty one figures. Ultimately, an academic is just a producer of prose. But as the pin factory needs the steelworks, or (only economists will get this one) the producer of blue coconuts needs the producer of red coconuts, so I am unable to say that every word in this thesis is wholly my own.

I cannot help but point out the irony that the most-read part of a dissertation—an *original* contribution to scientific knowledge, at least in theory—tends to be one long cliché. A speech at the Oscars, where you even have to do all the work by reading it yourself. And like at the Oscars, the only that thing *really* matters is the after-party.

The genre dictates mentioning that the dissertation would not be possible without the acknowledged persons. But in our advanced capitalist society, with its extreme division of labor, the list of people without whom this thesis would have been impossible is too long to mention. Certainly, it includes the Chinese factory workers that made my laptop, the farmers that grew my food and the developers of Stata.

Those who know me might think that I point out these unnamed soldiers just to be ironic, or even to trivialize the whole concept of an acknowledgment. In fact, my goal is just the opposite. While in a capitalist society, innumerable people are required to produce anything, most are paid to do so. However, some of the most important things in life are not provided by the market. It is this distinction that I want to point out: it is not that the following people necessarily contributed more to this dissertation than the developers of Stata (although some of them did), but because they did so out of a purely selfless motive, or at the very least because they went above and beyond what was required of them. *This*, I think, is what requires gratitude.

First, I should thank my promotor José Luis. Thank you for investing so much effort in me. Our cooperation included some hurdles. We started with a failed NWO

application and on multiple occasions failed to get the data we needed to move forward. Nevertheless, “JL” was always optimistic and helped me find a way forward. I also want to share my gratitude for showing me how the sausage is made: what do you need to do to become a successful researcher, apart from writing papers? His guidance and support, during the length of my PhD but particularly on the academic job market, have truly helped me to grow as a researcher and person.

I also want to thank my copromotor Mauro, in particular for exerting significant effort—including taking Dutch banks to court!—in gathering data that I could use for my thesis. While some of this data gathering took so much effort that the fruits came too late for this dissertation (the infamous *ar058* variable), Mauro deserves great gratitude not just from me, but from all of Dutch society for unlocking DNB’s data vault for scientific research.

Moreover, I want to thank Victor Aguirregabiria for hosting my visit at the University of Toronto, which changed the direction of my thesis profoundly. And of course, I want to thank the members of the doctoral committee—Bas van der Klaauw, Jaap Abbring, Alessandro Gavazza, Bertel Schjerning and Andrew Rhodes—for reading the thesis. While I have not been able to address all their comments in this dissertation, they have already helped tremendously while further developing the chapters contained therein as research articles.

I am also grateful to my colleagues at the VU Department of Economics. At the risk of forgetting somebody, I want to thank my fellow graduate students—Elisabeth, Casper, Cindy, Iulian, Zichen, both Davids, Diana, Bo, Yeorim, Philipp, Coen, Hermon, Rik, Luca, Benjamin, Martin, Florian, Piotr, Paul, Nadine, Sandor, Xiaoming, Simas, Lennart, Nahom, Lisette, Uwe—for making the slug more manageable. I also want to thank the faculty of Department for creating a stimulating work environment. I want to thank in particular Pieter, for lunch discussions on music and football, as well as Bas, Roland and Sabien, for having me as their teaching assistant and showing me what great teaching looks like. I also want to thank Pienke, who is probably the only person at the department whose absence would cause unavoidable chaos. Similarly, I want to thank every one at the Tinbergen Institute for my time there. Sometimes it was tough, but the people there—Albert Jan, Stijn, Andrej, Stephan, Paul, Aysil, Gavin, Gabrielle, Silvia, Francine, Agnieszka, Malin—made it worth it. Also many thanks to Ester and Christina for being invaluable on the job market.

This thesis would not have been possible without my friends and family. Thanks to the members of the “Spelletjesgroep”—Ruud, Josse, Mark, Micha and Reinier—for providing me with some much-needed diversion during the last years, and to Roel for sharing with me some unforgettable concert and festival experiences. I also thank to

thank my oldest friends—Mathieu, Steven, Florian, Jane, Joshua, Joyce, Leonie, Leonne, Thomas and Ruben. I think it is truly special that, even though we tend to be scattered all around the globe, we keep seeing each other. Particular thanks go out to Steven and Mathieu for the “weekly” luns at the VU, which we all had different reasons of periodically attending over the past few years. (I believe I am the first one to leave.) Because of this I think it is only fitting that they are my paranympths. Also many thanks to my family in Toronto—Brent, Sharon, David, Evan, Paul and of course Shelley—for hosting me during my time there. Finally, I would like thank my immediate family. While the last few years have not been the best times for us, I am grateful for the encouragement I have always received to follow my own path.

I can truly say that I without all you this thesis would not have been possible. And what do you get in return? This little book.

I hope you enjoy reading it.

Copenhagen, October 2019

List of Figures

1.1	Interest rates for new customers and renewers	14
1.2	Estimation strategy	17
1.3	Illustration of belief formation in SMPE	27
1.4	Marginal costs across banks	39
1.5	Counterfactual and observed interest rates	40
1.6	Counterfactual and observed interest rates when the Lasso is not used	41
2.1	Concentration in the Dutch mortgage market	64
2.2	Interest rates over time, by treatment and control market	71
2.3	The effect of entry for small and large incumbents	73
2.4	The response to entry by firm size	75
2.5	Estimated treatment effects over time	77
2.6	Coefficient of variation of monthly mortgage sales by incumbent	79
3.1	Welfare under commission-based and fee-based advice	96

List of Tables

1.1	Comparison of mortgages with and without government insurance (NHG)	11
1.2	Distribution of fixed interest period durations	12
1.3	Loan characteristics	13
1.4	Distribution of payment methods	14
1.5	Average interest rate difference between renewing and new customers . . .	15
1.6	Structure of banks' policy functions	36
1.7	Demand model estimates	37
1.8	Average marginal costs by payment method and fixed interest rate duration	38
1.9	The effect of the ban on history-based price discrimination on consumer surplus	41
1.10	The effect of the ban on history-based price discrimination on bank profits .	42
1.11	The interest rate difference between annuity and bullet mortgages, pre- and post-ban	44
1.12	Predictions of pre-ban interest rates from the post-ban model	46
1.13	Consistency of estimated policy functions with sparse maximization	50
1.14	Comparison of the estimated effect of the ban on history-based price discrimination for different methods of choosing the Lasso penalty parameter . . .	51
2.1	Parameter values for which the model is simulated	62
2.2	Entry in the Dutch mortgage market in 2014	64
2.3	Market shares (%) of payment types	67
2.4	Switching between treatment and control markets	69
2.5	Main results: the effect of entry on incumbents' interest rates	74
2.6	Borrower and loan characteristics for the treatment and control market . . .	78

Introduction

Since its inception, a central question in the field of industrial organization has been the effect of firm conduct on market performance (Robinson 1933; Bain 1968). A classical question is, for example, whether it matters whether firms compete in prices or in quantities (Kreps and Scheinkman 1983). Many such questions are *static* in nature. This nature can manifest itself in two ways. The first is that some of firms' choices, such as investment, are considered sunk and hence fixed. The second is that choices firms make today, for example about their price, have no effects tomorrow. In either case, it is assumed that time in some sense plays no role.

Some of the more interesting questions in industrial organization are however inherently *dynamic*. For example, a static analysis holds the number of firms active in a market fixed. However, firms might choose to enter or exit the market over time (Ericson and Pakes 1995). Moreover, many of the most interesting strategies firms employ are dynamic as well. For example, a firm might charge low prices today to expand its market share, in the hope that it reaps the profits from this strategy tomorrow (Klemperer 1987b).

This dissertation considers three questions of such a dynamic nature, all in the context of the Dutch mortgage market. All chapters consider a change to the market, which I use to say something about firms' dynamic conduct. In other words, in this dissertation I go "back to the future" to better understand firm behavior and its effects. (All chapters are single-authored.)

Chapters 1 and 2 consider markets with switching costs. Markets have switching costs when it is costly for consumers to purchase from a different firm than their current supplier, either due to contractual costs or due to the opportunity cost of time. While there is a sizable literature that studies the effects of the level of switching costs, either theoretically or empirically (Farrell and Klemperer 2007, see e.g. Shcherbakov 2016),

there is little empirical work that examines firms' strategies in markets with switching costs.

Chapter 1 considers a common strategy in markets with switching costs: the tendency to charge different prices to new than to renewing customers, a practice called history-based price discrimination. Theoretically, it is not obvious whether this practice is preferable to firms charging a single, uniform price to both consumer groups. Therefore, in this chapter I evaluate a regulation in the Dutch mortgage market that prohibited banks from engaging in history-based price discrimination. Comparing the market with and without this regulation then allows me to say something about the effects of history-based price discrimination.

A major technical challenge in this analysis comes from the dynamic nature of banks' profit maximization problem. When there are switching costs, it matters how many mortgages a bank has sold in the past—the more it has sold, the more “locked-in” renewing customers it has, and the higher the interest rate it can charge. Hence, banks need to keep track of many variables to determine their optimal interest rates. It turns out that the researcher then needs to perform calculations for every possible combination of these variables. The number of required calculations quickly increases in the number of variables banks keep track of. For example, if there are two variables, each taking ten different values, $10^2 = 100$ calculations are required, while if there are five, this number increases to $10^5 = 10000$! This *curse of dimensionality* is a major impediment to empirical analyses of questions of a dynamic nature.

To deal with this problem, I introduce a new method for estimating dynamic games, Sparse Markov Perfect Equilibrium (SMPE). In an SMPE, agents display rational inattention towards the state: they react to changes in some state variables but not to changes in others (Gabaix 2014). This reduces the number of relevant states for the researcher as well, and as it turns out, it is easy to identify which state variables agents pay attention to using standard techniques from machine learning (Tibshirani 1996).

The main results of this chapter are as follows. I find that the prohibition of history-based price discrimination lead to an average welfare increase of €125 per mortgage per year. Consumer surplus also increased, by €415 per mortgage per year. Bank profits on the whole decreased, by €290 per mortgage per year on average.

Chapter 2 studies the way banks compete in the Dutch mortgage market *after* history-based price discrimination has been banned. Theory, e.g. Klemperer (1987b) and Beggs and Klemperer (1992), predicts that in markets with switching costs, firms face a trade-off between charging a low price to attract customers today (“investment”) and charging a high price to exploit existing locked-in customers (“harvesting”). However, evidence that firms in fact behave this way is scant. In this chapter, I provide evidence that

firms in fact price according to this investment-harvesting trade-off. I show using the model of Somaini and Einav (2013) that after entry, these dynamic pricing incentives are amplified. To be precise, firms with fewer locked-in customers should decrease their price more after entry than firms with more locked-in customers.

I test these predictions using recent entry by pension funds into the Dutch mortgage market. These entrants do not sell all types of mortgages available in the market. Hence, the entry splits the Dutch mortgage market into a treatment market, consisting of mortgages sold by both incumbents and entrants, and a control market, consisting of mortgages sold only by incumbents. I can then use a simple difference-in-difference framework to test the model's predictions.

I find that smaller banks indeed decreased their interest rates more after pension funds entered the market than larger banks. In this sense, I provide empirical evidence for the investment-harvesting trade-off. Zooming in, I find that past market shares can explain about 12% of post-entry interest rate responses. Hence, it seems that *other* mechanisms might also be at work in markets with switching costs. I leave to further research what these mechanisms might be.

Chapter 3 is about the role of advisers in financial markets (including, but not limited to, the Dutch mortgage market). Historically, such advisers are paid through commissions or kickbacks by upstream firms. A common concern about such payments is that they cause advisers to be *biased*, that is, that they have an incentive to recommend a product with a high kickback rather than the product that is most suitable for the consumer. As a result, some regulators, including the Dutch AFM, have banned such kickbacks. In the absence of kickbacks, advisers receive a fixed fee paid by consumers for their services.

While a static analysis of this regulation (e.g. Inderst and Ottaviani 2012b) might predict that it is beneficial for consumers because it removes advisers' conflict of interest, the first real-world experiences with commission bans have not been uniformly positive. A major concern for regulators is that advisers leave the market after commissions are banned, creating a so-called "advice gap" in which advice is no longer accessible for all consumers. Hence, it seems that dynamic concerns are also important in these kind of markets.

To study these concerns, I develop a theoretical model of advice in which, contrary to the existing literature, advisers' presence is endogenous. I show that endogenous entry and exit can be crucial to understand the effects of commission bans: while in a static setting banning kickbacks is beneficial for consumers, this is not the case when allowing advisers to enter or exit the market. The reason is that competition drives the fees that advisers can charge down. As a result, adviser profitability decreases after

a commission ban. Therefore, advisers exit the market and the accessibility of advice deteriorates. Hence, this chapter shows the importance of considering the potential endogeneity of market structure when regulating advice.

Price discrimination, switching costs and welfare: Evidence from the Dutch mortgage market

1.1 Introduction

In many markets with switching costs firms charge a lower price to new customers than to existing customers. This phenomenon, called history-based or behavior-based price discrimination, has been documented for credit markets (Ioannidou and Ongena 2010; Barone, Felici, and Pagnini 2011), cellular contracts (Alé 2013) and newspapers (Asplund, Eriksson, and Strand 2008). Whether history-based price discrimination increases or decreases welfare is theoretically ambiguous. Compared to uniform pricing, firms have an incentive to charge relatively high prices to renewing customers, because they face switching costs. On the other hand, they want to charge relatively low prices to new customers to entice them to switch. Since either effect can dominate, the effect of history-based price discrimination on welfare is an empirical question. Despite the fact that history-based price discrimination is common, this question has so far remained unanswered.

This paper exploits a natural experiment in the Dutch mortgage market to estimate the effects of history-based price discrimination on consumer surplus, bank profits and welfare. To do so, I develop a structural model of demand and supply of this market. The supply side of the model is a dynamic game and suffers from the curse of dimensionality. To estimate the model, I therefore introduce a new method to estimate dynamic games with large state spaces. The method is based around a new solution concept for dynamic games, Sparse Markov Perfect Equilibrium, in which firms display

partial attention to the state. I show that estimation techniques from the machine learning literature can identify the variables firms pay attention to.

In the Dutch mortgage market, most households fix their interest rate for a certain period, most commonly ten years. When this period ends, they can either renew their mortgage at their current bank or switch to a different one. However, switching is costly: for the average household, switching costs are around €3500 in addition to the opportunity cost of time.¹ Therefore, banks have an incentive to charge high interest rates to existing customers, since they are “locked in” because of switching costs. This is called the *rent extraction effect*. On the other hand, poaching customers from other banks becomes more attractive compared to uniform pricing because history-based price discrimination allows banks to charge them lower interest rates without cannibalizing the profits on their captive customers. This downward pressure on interest rates is called the *competition effect*. Depending on whether the rent-extraction or the competition effect dominates, the average interest rate can be higher or lower under history-based price discrimination than under uniform pricing. Moreover, because the price difference between a consumer’s current firm and its competitors is larger than under uniform pricing, history-based price discrimination encourages switching. When switching is costly, this is socially wasteful. Finally, cross-segment inefficiencies may occur when a consumer purchases from an inefficient bank because it obtains a relatively low interest rate there (Stole 2007). For these reasons, the question whether history-based price discrimination increases or decreases consumer surplus, profits and total welfare can only be answered empirically.²

Consistent with this theory, prior to 2013, Dutch banks offered higher interest rates to renewing customers than to new customers. Regulators were concerned that such history-based price discrimination was harmful to consumers. As a result, regulations were introduced stipulating that from 2013 onwards banks have to offer the same interest rate to new and to renewing customers if those customers have a similar risk profile. I exploit this ban to study the effects of history-based price discrimination on consumer surplus, bank profits and total welfare. I do so using administrative data from the Dutch central bank, which include the universe of mortgages from institutions

¹Typical costs include advice, taxation, notary and insurance fees. A calculation by the country’s largest mortgage broker suggests total costs of about €6000 for an average household. Since these costs are tax deductible and the average marginal income tax rate is 42%, the average household will incur monetary costs of about €3500. (<https://www.hypothekeer.nl/jouw-woonsituatie/hypotheek-oversluiten/>, Accessed March 8, 2017.)

²The theoretical literature on history-based price discrimination in the presence of switching costs (Chen 1997; Taylor 2003; Gehrig, Shy, and Stenbacka 2012; Rhodes 2012) generally finds that, compared to uniform pricing, history-based price discrimination can both increase and decrease consumer welfare, and is divided on the predicted effect on firm profits and total welfare.

under its supervision.³ To rule out as much as possible that interest rate differences between customers are caused by differences in risk, I focus on mortgages that are insured by the Dutch government.

I find that before the 2013 ban there were indeed significant interest rate differences between renewing and new customers. On average, a renewing household paid an interest rate that was .32 percentage points higher than a new customer. Since the average mortgage in my sample is about €150,000 this means that renewing households paid around €278 per year more in interest (after tax deductions).

To assess the effects of the ban on history-based price discrimination, I estimate a structural model of demand and supply of the Dutch mortgage market. Because higher sales today imply more locked-in consumers in the future, banks play a dynamic game. On the demand side, I allow for rich interactions between household and product characteristics in consumers' utility specification, since the curvature of the demand function is a crucial determinant of the effect of third-degree price discrimination on welfare (Holmes 1989). This demand-side heterogeneity implies that the state space of the dynamic game that banks play is very large: the pay-off relevant state contains the full joint density of previous market shares and household demographics. Traditional methods for estimation of dynamic games cannot deal with games with such large state spaces.⁴ Two solutions to this problem have been used in the literature. One is to restrict the amount of heterogeneity in the model, for example by having only a small number of types of consumers in the market.⁵ This is not an attractive option when studying price discrimination, since this limits the shapes the demand function can take. A second solution is to make ad hoc assumptions to reduce the dimension of the state space.⁶ However, there are typically many different ad hoc assumptions one could make. Moreover, I show that, at least for my application, some on the face reasonable ad hoc assumptions lead to very strange results.

³These institutions have a combined market share of 75% - 80% (Mastrogiacomo and Van der Molen 2015).

⁴Most methods, most prominently Bajari, Benkard, and Levin (2007), contain a first stage in which firms' actions are regressed on state variables. If the number of state variables is larger than the number of observations, this is impossible. So-called nested fixed point methods, in which an equilibrium is calculated for every candidate parameter vector, take too much time when the state space is large. The nested fixed point method of Abbring et al. (2017) is fast, but only applicable to models of firm entry and exit.

⁵One recent example of this approach is Cosguner, Chan, and Seetharaman (2016).

⁶For example, many papers group states together. A prominent example is Collard-Wexler (2013), who groups plants of different sizes together and ignores markets with too many firms to reduce the state space. Another common strategy is to assume that firms' behavior only depends on some average or total value of the state instead of the full distribution. For example, Kalouptsi (2014), assumes that ship manufacturers' value functions depend only on the total backlog in the market and Barwick and Pathak (2015) assume that real estate agent's commissions depend only on average housing market conditions. Another strategy is to assume the policy functions take a particular form (Wollmann 2018).

I instead solve the curse of dimensionality by introducing a new method for estimating dynamic games. The method combines economic theory with techniques from machine learning. On the theory side, I dispense with the assumption that firms play a Markov Perfect Equilibrium (MPE). Instead, I introduce a new solution concept that is more amenable to estimation: Sparse Markov Perfect Equilibrium (SMPE). An SMPE is an MPE in which firms perform sparse maximization. Under sparse maximization, which was introduced by Gabaix (2014) and which I extend to dynamic games, firms optimally condition their policy functions only on a subset of the payoff-relevant state variables. This can be motivated by relaxing the usual implicit assumption that firms' information on the state is free. The relevant state in an SMPE is smaller than in an MPE so that an SMPE can be estimated in situations in which estimating an MPE is impossible. Which state variables firms pay attention to can be estimated using variable selection techniques from the machine learning literature.

Econometrically, I use a two-stage method as in Bajari, Benkard, and Levin (2007), where I use the Lasso in the first stage to estimate the policy functions. The Lasso selects the state variables firms pay attention to. For a given estimated model, SMPE imposes testable restrictions on the data. These restrictions can be used to test whether the econometric selection of state variables has a sensible economic interpretation. Thus, it is the interplay between theory and econometrics which allows me to deal with the curse of dimensionality.

I find that the ban on history-based price discrimination leads to economically significant increases in welfare, of about €125 per mortgage or €11.5 million in total per year. The ban causes a drop in interest rates, so that they are closer to marginal costs. Moreover, the ban leads to less socially wasteful switching. Finally, there is a reallocation effect: the ban causes the market share of low-cost banks to increase. The welfare increase accrues more than completely to consumers: consumer surplus increases with about €415 per year for an average mortgage. On the flip side, the average profit per mortgage decreases by €290.

This paper makes several contributions. First, I am, to the best of my knowledge, the first to empirically study whether history-based price discrimination increases or decreases welfare. The only other empirical study on history-based price discrimination in markets with switching costs is Cosguner, Chan, and Seetharaman (2017), who only consider the effect on profits. In particular, I find that cost asymmetries between firms can be of first-order importance for the welfare effects of history-based price discrimination. Although the importance of asymmetries has sometimes been recognized in the literature on price discrimination generally (e.g. Stole (2007)), the theoretical literature on history-based price discrimination has focused only on symmetric cost functions.

Therefore, I show that this assumption is not innocuous and that cost asymmetries can be very important. This mirrors the recent finding of Asker, Collard-Wexler, and De Loecker (2017), who find that cost asymmetries are crucial to understand the effects of market power.

The second contribution of this paper is that it provides a tractable empirical model of supply for markets with switching costs. Switching costs, or demand inertia more broadly, have been documented in many markets.⁷ These include markets that are highly policy-relevant, such as health care (Nosal 2012; Handel 2013). Demand inertia also plays a key role in many macroeconomic models with so-called “customer markets” (Phelps and Winter 1970; Bils 1989), and was recently found to play an important role in inflation dynamics (Gilchrist et al. 2017). To assess potential policy interventions in markets with switching costs, it is important to understand how firms might respond to such measures. Some previous work, which I discuss in more detail below, has developed empirical models of firm behavior in markets with switching costs. However, these models use certain assumptions that limit their applicability. The method I introduce in this paper, SMPE, is generally applicable and can (contrary to previous models) accommodate rich heterogeneity, non-anonymous strategies and a large number of firms.

The final contribution of this paper is to provide a new method to estimate dynamic games with large state spaces. Here, I contribute to the literature on the estimation of dynamic games (Aguirregabiria and Mira 2007; Bajari, Benkard, and Levin 2007; Pakes, Ostrovsky, and Berry 2007; Pesendorfer and Schmidt-Dengler 2008). My method is most closely related to Bajari, Benkard, and Levin (2007). Like them, I follow a two-step method, where in the first step firms’ policy functions are regressed on state variables and in a second step the deep parameters of the model (typically marginal costs) are recovered. My main innovation is to use the Lasso in the first step. For this, I also give a micro-foundation in the form of SMPE.

The remainder of this paper is organized as follows. Section 2 discusses the related literature in more detail. Section 3 gives some background on the Dutch mortgage market and its ban on history-based price discrimination. Section 4 describes my data set and gives reduced-form evidence on the existence of history-based price discrimination in the Dutch mortgage market. Section 5 describes the structural model of demand and supply that I estimate. It also introduces the concept of Sparse Markov Perfect Equilibrium. Section 6 discusses identification and estimation of my model. Section 7 gives the estimation results, as well as an assessment of the welfare effects of the ban on history-based price discrimination. Section 8 discusses the robustness of my model,

⁷For an overview of papers estimating switching costs, see the literature review below.

while Section 9 concludes.

1.2 Related literature

In addition to the literature summarized in the introduction, this paper is broadly related to two strands of literature: one on history-based price discrimination and empirical models of state-dependent demand and the other on simplifying solution concepts for dynamic games.

While there is a relatively large literature documenting the existence of history-based discrimination (Asplund, Eriksson, and Strand 2008; Ioannidou and Ongena 2010; Barone, Felici, and Pagnini 2011; Alé 2013), there is not much work estimating its effects. The aforementioned Cosguner, Chan, and Seetharaman (2017) is an exception. They use counterfactual simulations to study whether history-based price discrimination would be profitable in the cola industry. They however do not have the exogenous variation in pricing (with and without history-based price discrimination) that I have. In addition, I also look at consumer surplus and total welfare and not just at profits.

My paper is related to various other papers that estimate dynamic models of firm pricing in the presence of state-dependent demand. Fleitas (2017) develops a dynamic model of insurer pricing in Medicare Part D. His model restricts strategies to be anonymous and symmetric and does not allow for rich consumer heterogeneity in the demand model; my model has neither restrictions. The anonymity assumption is particularly restrictive when studying oligopolies: in such markets it is difficult to imagine that the identity of the firm that consumers purchase from is irrelevant. Cosguner, Chan, and Seetharaman (2016) develop an empirical model of supply in the presence of switching costs of the cola market. Their estimator, based on forward iteration of the value function, is feasible because their market features only two firms. As the number of states for which forward iteration of the value functions must be performed increases exponentially in the number of firms, their estimator would quickly become infeasible for markets with more firms. My approach can (and, in this paper, does) handle a larger amount of firms. MacKay and Remer (2018) develop a model of supply with demand inertia in the context of gasoline markets. They require observations on firms' marginal costs to estimate their model, my approach does not. Rickert (2016) studies the German diaper market: his model has a finite horizon, which is not an attractive assumption in many markets. None of these papers feature markets with history-based price discrimination.

More broadly, my paper contributes to a growing literature documenting and estimating switching costs (Viard 2007; Dubé, Hitsch, and Rossi 2009; Cullen and

Shcherbakov 2010; Miller and Yeo 2012; Nosal 2012; Handel 2013; Honka 2014; Cullen, Schutz, and Shcherbakov 2015; Ho 2015; Shcherbakov 2016; Raval and Rosenbaum, Forthcoming; Weiergräber 2017). This literature estimates the effects of (different levels of) switching costs on market outcomes. I take the level of switching costs as given and study a potential consequence of switching costs, namely history-based price discrimination.

Secondly, my paper contributes to a literature on simplifying solution concepts for dynamic games. In recent years, various solution concepts have been introduced, such as oblivious equilibrium (Weintraub, Benkard, and Van Roy 2008), experience based equilibrium (Fershtman and Pakes 2012) and moment-based Markov equilibrium (Ifrach and Weintraub 2017). These approaches make *computation* of equilibria of dynamic games easier. The notion of SMPE that I introduce makes *estimation* easier. SMPE is closest to the moment-based Markov equilibrium of Ifrach and Weintraub (2017). Ifrach and Weintraub (2017) make the assumption that firms pay attention to the full state of dominant firms and some moments of the state of fringe firms. In an SMPE, no such assumptions are necessary. Instead, firms choose which variables they pay attention to based on a cost-benefit analysis. This means that SMPE can also be used when the dimension of a common state or of the states of dominant firms is large.

1.3 The Dutch mortgage market

Because the Dutch mortgage is in some aspects quite different from mortgage markets in other countries, I begin with a brief overview of its most important features. In the Netherlands, mortgages are primarily sold by banks. The market is reasonably concentrated, with an HHI of 2100. Three banks (ABN Amro, ING and Rabobank) dominate the market, with a competitive fringe consisting of smaller banks and pension funds. Approximately 55% of households own their house.⁸

In the Dutch mortgage market, many different types of mortgages are sold. Two categories can be distinguished. The first consists of non-amortizing mortgages—mortgages where the principal is paid in a lump sum at the end date. Such mortgages include bullet, savings, life and investment mortgages; the latter three are sold together with a financial product, the returns of which are used to pay off the principal at the mortgage's end date. The second type of mortgage are amortizing mortgages: for these regular payments towards the principal are made. Amortizing mortgages, which are more

⁸Statistics Netherlands. <http://statline.cbs.nl/StatWeb/publication/?VW=T&DM=SLNL&PA=71446ned>. Accessed January 31, 2017.

common in most other countries, exist in the form of annuity and linear mortgages.⁹

Mortgages in the Netherlands are, contrary to what is typical in the United States, with recourse so that consumers are personally liable for any outstanding mortgage debt in case of default. However, mortgages smaller than the average national house price are typically insured by the government through the so-called national mortgage guarantee (Nationale Hypotheek Garantie, or *NHG*, in Dutch). The NHG pays off the remaining balance if a household defaults on its mortgage because of divorce, disability or unemployment.¹⁰ Enrollment in the NHG costs 1% of the loan sum. Banks view mortgages with NHG as low risk and offer significant interest rate discounts if consumers choose to enroll.¹¹

Households tend to fix their interest rate for relatively long periods, most commonly for ten years (Table 1.2). However, this fixed interest rate period is shorter than the typical duration of a mortgage, which is thirty years. When the fixed interest rate period ends but the principal is not yet due, a household's current bank offers it to renew its mortgage. At this point, however, it is also possible to switch to a different bank.¹² However, switching is costly. Switching costs typically include the costs of a notary, the cost of appraisal, and, if the mortgage qualifies for NHG insurance, insurance fees. The country's largest mortgage broker estimates that these costs are around €3500 for an average household, accounting for the fact that these costs are tax-deductible.¹³

History-based price discrimination, of which I give evidence below, existed in two forms. Some banks would explicitly have a different interest rate for renewing customers. Other banks, however, applied price discrimination in more implicit ways, through discounts that in practice were only available to new customers.¹⁴ One common example of the latter is a discount for first-time buyers: this discount typically was no longer available upon renewal.

1.3.1 Ban on history-based price discrimination

The origin of the Dutch ban on history-based price discrimination lies in the observation that after a large initial drop during the 2008 financial crisis, mortgage interest rates

⁹An annuity mortgage features constant mortgage payments. A linear mortgage, on the other hand, features constant amortization, so that mortgage payments are decreasing over time.

¹⁰NHG. <https://www.nhg.nl/Consument/Wat-is-NHG>. Accessed January 31, 2017.

¹¹A typical discount is between .4 and .7 percentage points (Fransman 2017).

¹²Switching to a different bank is also possible when the fixed interest rate period is still ongoing. However, a household incurs severe penalties when it ends the fixed interest rate period prematurely.

¹³See footnote 1 for the source and further explanation.

¹⁴Dutch Competition Authority (in Dutch), pp. 29-30. "Sectorstudie Hypotheekmarkt." www.acm.nl/sites/default/files/old_publication/bijlagen/7091_Sectorstudie%20Hypotheekmarkt.pdf. Accessed September 19, 2018.

in the Netherlands quickly increased again from 2009 onwards, while they stayed low in other European countries (Dijkstra, Randag, and Schinkel 2014b). Various possible reasons have been given for this, from collusion to the large reliance of Dutch banks on external funding, the cost of which increased sharply after the financial crisis.¹⁵ After an investigation, the Dutch competition authorities saw no reason to intervene, but they did provide some recommendations to make the Dutch mortgage market more competitive (Nederlandse Mededingingsautoriteit 2011). One recommendation, which the government followed, was to ban history-based price discrimination, as the competition authorities believed the differences in interest rates between prolonging and first-time customers to be anti-competitive. As a result, the ban on history-based price discrimination came into effect on January 1, 2013.

The regulation states that institutions are legally obliged to offer the same interest rate to households at the end of their fixed interest rate period as to households who are first-time customers at that institution, if those two households have a similar risk profile. As the responsible regulator AFM later clarified, this does not mean that banks are barred from all types of price discrimination (Autoriteit Financiële Markten 2015). For example, they are allowed to (and in practice do) offer a discount if a household also has a deposit account at the same institution. However, such discounts have to be equally available to existing and first-time clients.

1.4 Data

This section describes my data. I begin by introducing my data set. Then I explain how I construct my sample and give some descriptive statistics. Finally, I show reduced-form evidence of history-based price discrimination in the Dutch mortgage market.

1.4.1 Data sources

My main data source is the Loan-Level Data (LLD) from the Dutch Central Bank (DNB). The LLD are a yearly panel containing micro-level data covering almost the complete Dutch mortgage market. Starting in 2013, participating banks hand in a yearly report of all their outstanding mortgages. The LLD contain detailed information on the loans, as well as some information on the underlying property and the household that purchased it. Not all banks are required to submit information to the LLD: small banks and some foreign banks do not have to report information. The LLD cover between 75% and

¹⁵See Dijkstra, Randag, and Schinkel (2014a) and the other articles in the same issue of the *Journal of Competition Law and Economics* for an overview of the arguments.

80% of the full market and aggregate statistics match those from other data sources (Mastrogiacomo and Van der Molen 2015). The banks in my sample are ABN Amro, Florius, ING, Obvion and Rabobank, as well as some fringe players which I aggregate into a single bank. The only bank with a significant market share that is missing is SNS Reaal, which does not provide information on whether or when a mortgage is renewed.¹⁶ To preserve banks' anonymity—a pre-condition for accessing the LLD—I will anonymize bank names for the remainder of this paper.¹⁷

This study is based on the LLD from 2013, 2014 and 2015, so that I have three years of observations. However, since I observe the *stock* of all mortgages in 2013 and most mortgages have a fixed interest rate period of at least five years, I can see virtually all purchases of the directly preceding years.

In addition to the LLD, I use the DNB Household Survey (DHS). The DHS is a yearly survey of a random sample of approximately 2000 Dutch households on their finances. I use the DHS because the LLD are a choice-based sample: it does not provide any information on households that purchase no mortgage. To say something about the extensive margin of the market, I add the demographic information on those households from the DHS. Since the LLD are a choice-based sample and the DHS is a random sample, I adjust for different sampling probabilities in the demand estimation.

1.4.2 Inferring switching

Because of the panel dimension of my data, I can observe a household's previous mortgage when it purchases a new mortgage, allowing me to observe whether it renewed its current mortgage or switched to a different bank. However, the panel dimension of my data is limited by the fact that banks use different schemes to encode household id's. This means that when a household switches to another bank, I cannot observe to which bank exactly. To address this issue, I probabilistically match switching households based on the birth year of the primary borrower, the type of mortgage (e.g. bullet or annuity), the outstanding balance of the mortgage at the moment of switching and the maturity year of the loan. Further details can be found in Appendix 1.B.1.

Since the first wave of the LLD is 2013, I do not have a "previous" observation for households that make a purchasing decision before 2013. I *can* observe whether or not the mortgage is renewed or not as there is a field that indicates this. Therefore, before 2013, I know the previous bank of renewing households—as it equals their current

¹⁶SNS Reaal has a market share of approximately 8%. This is significantly smaller than the market shares of ABN Amro, ING and Rabobank, who all have a market share of more than 20%.

¹⁷For this reason, I am also unable to provide market shares of the anonymized banks, because the market shares allow identification of the banks based on publicly available data.

Table 1.1: Comparison of mortgages with and without government insurance (NHG)

	NHG	No NHG
Age household head (years)	38.0	49.0
Household income (€)	44730	70537
Initial loan (€)	141668	113502
Property valuation (€)	188254	330344
Observations	436608	229621

Note: The table compares average household and loan characteristics for loans with and without NHG insurance, for mortgages with a fixed interest rate period starting in the period 2010-2015.

bank—but not of switching households or households that purchase their first mortgage. I deal with this by estimating the model on the part of the data for which I can identify switching behavior well, i.e. for the post-2013 data, as I explain in more detail below.

1.4.3 Sample selection

My empirical strategy depends on comparing the interest rate that new and renewing households pay for the same mortgage. One reason other than purchasing history for such interest rate differences may be differences in risk. To rule out as much as possible that interest rate differences are caused by differences in risk, I focus on mortgages with NHG insurance. As explained in Section 1.3, such mortgages are insured by the government so that from the perspective of banks they are more or less risk free. However, mortgages are only eligible for NHG insurance when the purchase amount is at most €245,000.¹⁸ As a result, households that have a mortgage with NHG differ systematically from those that do not. Table 1.1 shows that on average, households with NHG are 11.5 years younger, their household income is €15,000 lower and they own properties valued €150,000 less than households without NHG. However, households with NHG mortgages have larger loans because households who do not enroll into the NHG scheme tend to make larger down payments. My results should thus be understood to be applicable to this segment of the overall mortgage market.

I further restrict my sample by only considering mortgages with a fixed interest rate period of around five, ten, fifteen or twenty years. If a mortgage has a fixed interest rate period within six months of any of these periods, I round off the duration to that length. For example, I round a fixed interest rate period of 57 months to six years. The reason for not considering other fixed interest rate periods is that they all have very small market shares.

¹⁸In 2015. The maximum purchase amount was somewhat higher in previous years, up to €276,190 in 2012.

1. PRICE DISCRIMINATION, SWITCHING COSTS AND WELFARE

Table 1.2: Distribution of fixed interest period durations

	2010	2011	2012	2013	2014	2015
5 years	11.29	22.16	21.82	27.15	21.07	8.46
10 years	81.12	69.14	66.96	67.71	68.73	65.87
15 years	1.96	3.12	3.60	2.60	3.50	7.18
20 years	5.63	5.58	7.63	2.54	6.70	18.49
Observations	73923	69363	64721	65843	78240	84809

Note: The table shows the proportion (%) of fixed interest rate durations by start year of the fixed interest rate period. Fixed interest rate durations within six months of one the stated durations are counted towards the market share of that duration. Mortgages with other fixed interest rate durations are discarded.

In my data there remain interest rate differences between households purchasing the same type of loan in the same month even after controlling for history-based price discrimination. These differences are caused by the fact that I observe the date on which the loan deed was signed, which is different from the date the interest rate was offered: e.g. one household who signed their mortgage in March might have received their interest rate offer in January and another their (different) offer in February. A second reason for these differences is the existence of other types of price discrimination which I do not observe. Most banks for example offer a small discount on the interest rate if a household opens a deposit account at the same bank. Since I do not have information on the source of this residual price dispersion, I take as a mortgage's interest rate its modal interest rate in a given month, and assume that every household paid this interest rate. Some mortgages have zero sales in certain months. For those loans I impute the interest rate based on a simple linear regression.¹⁹

1.4.4 Descriptive statistics

Table 1.2 shows the market shares of the fixed interest rate durations of the loans in my sample. A fixed interest rate period of ten years is by far the most common, with a market share between 67% and 82%. A five year period is the second most common, with relatively small market shares for fifteen and twenty years. In 2015 the market share of twenty year fixed interest rate periods increases dramatically, probably because consumers wanted to "lock in" the historically low interest rates in 2015.

Table 1.3 describes the main characteristics of loans in my sample. The average interest rate decreases from 2010 (4.66%) until 2015 (2.77%). The average fixed interest

¹⁹I run separate regressions for the interest rates for new and renewing customers. I include bank, loan type, fixed interest rate period and month fixed effects. The regressions have an R^2 of .827 (new customer interest rate) and .886 (renewing customer interest rate).

Table 1.3: Loan characteristics

	2010	2011	2012	2013	2014	2015
Interest rate	4.66 (0.45)	4.66 (0.49)	4.49 (0.51)	4.11 (0.55)	3.47 (0.52)	2.77 (0.44)
Current balance (€)	114616 (51666)	116587 (52965)	114927 (50929)	114228 (52504)	116261 (54354)	118380 (51275)
Fixed interest rate period (months)	121 (36)	115 (41)	118 (45)	108 (36)	117 (43)	141 (53)
Renewed mortgage	0.099 (0.299)	0.132 (0.339)	0.106 (0.308)	0.233 (0.423)	0.350 (0.477)	0.398 (0.489)
Observations	73923	69363	64721	65843	78240	84809

Note: The table shows average characteristics of loans with NHG insurance by start year of the fixed interest rate period. Standard deviations are in parentheses.

rate period is approximately ten years, comparable to the mode. The proportion of loans that is renewed rather than a new purchase increases dramatically after the ban on history-based price discrimination in 2013, from around 10% before to between 23% and 40% after. This means that households switch less often to a different bank after their fixed interest rate period ends. The average switching rate in my sample is 6%.

Table 1.4 shows that before 2013, non-amortizing loans such as bullets and savings mortgages were by far the most popular. This is because, before 2013, the interest rate on mortgages was fully deductible, so that non-amortizing mortgages were very attractive. From 2013 onwards, new non-amortizing mortgages no longer qualify for interest rate deductibility. Therefore, the market share of non-amortizing loans becomes much smaller. Since this change in taxes happens at the same time as the ban on history-based price discrimination, it is a possible confounder. In the demand model, the tax change amounts to a change in households' choice set for which I can control.²⁰ It also possible that the tax change causes a change in banks' pricing. In Section 1.8.1 I show evidence that this is not the case.

1.4.5 Evidence of history-based price discrimination

In this section, I present reduced-form evidence on the extent of interest rate differences between new and existing customers and the effect of the ban on history-based price discrimination. Figure 1.1 plots the average interest rates paid by renewing and new

²⁰The reason the tax change can be interpreted as a change in choice sets is because the tax change made non-amortizing mortgages so unattractive that in practice banks do not even offer them any more to first-time buyers.

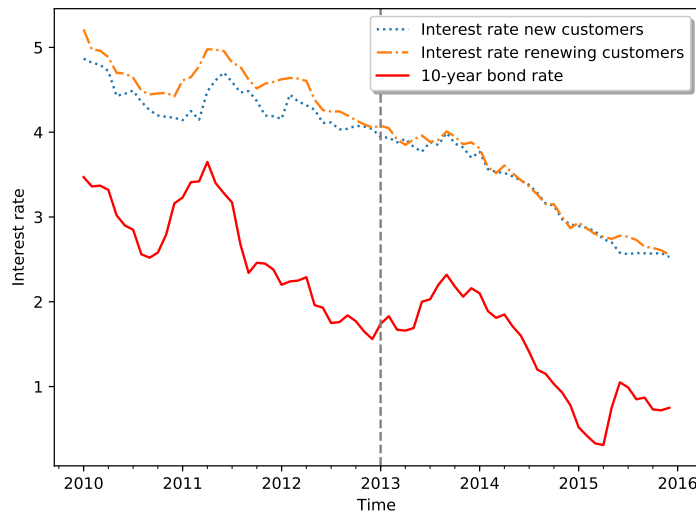
1. PRICE DISCRIMINATION, SWITCHING COSTS AND WELFARE

Table 1.4: Distribution of payment methods

	2010	2011	2012	2013	2014	2015
Annuity	2.58	3.29	5.96	33.40	48.41	43.81
Linear	0.37	0.45	0.68	2.34	4.25	4.47
Bullet	20.85	18.32	18.20	16.12	16.58	22.18
Savings mortgage	66.80	68.00	68.48	34.30	13.77	14.12
Life mortgage	7.41	8.15	5.32	11.70	14.36	12.39
Investment mortgage	1.99	1.78	1.36	2.14	2.64	3.03
Observations	73923	69363	64721	65843	78240	84809

Note: The table shows the distribution of payment types of loans with NHG insurance by start year of the fixed interest rate period.

Figure 1.1: Interest rates for new customers and renewers



Note: The figure shows the average interest rates for households with a NHG mortgage. Both interest rates are weighed by the total market share, i.e. the sum of the market shares across new and renewing customers. The interest rate on 10-year Dutch government bonds is included for comparison.

customers. Before the ban, renewing customers clearly pay higher interest rates. These differences largely disappear in 2013, after the ban.

To get at the magnitude of price discrimination, I regress the interest rate consumers pay on month by loan fixed effects and whether the loan is renewed or not. Thus, I compare the interest paid by renewing and new customers for the same loan in the same month.²¹ Table 1.5 shows that, before the 2013 ban, renewing consumers paid statistically higher interest rates than new consumers. This difference is also economically significant. Since the average mortgage's starting balance is around

²¹In Figure 1.1, I do not control for loan fixed effects. Therefore, differences in Figure 1.1 might be caused by renewing and new customers purchasing different mortgages.

Table 1.5: Average interest rate difference between renewing and new customers

	2010	2011	2012	2013	2014	2015
Renewed mortgage	0.328*** (0.0167)	0.401*** (0.0190)	0.263*** (0.0140)	0.0925*** (0.00886)	0.0806*** (0.00829)	0.193*** (0.0159)
Constant	4.631*** (0.00166)	4.604*** (0.00251)	4.461*** (0.00149)	4.089*** (0.00207)	3.438*** (0.00290)	2.698*** (0.00635)
Loan \times month fixed effects	✓	✓	✓	✓	✓	✓
Observations	73923	69363	64721	65843	78240	84809

Note: The table shows the average difference in interest rates between renewing and new customers for the same loan. A loan is defined by the originating bank, the length of the fixed interest rate duration and the payment method. Standard errors are in parentheses. *** $p < 0.001$.

€150,000, the differences in interest rates imply that renewing consumers paid between €228 (in 2012) and €348 (in 2011) after tax in yearly interest payments more for the same type of loan as new consumers. These estimates conform to previous estimates made by the AFM, the regulatory agency charged with upholding the ban, in the run-up to the 2013 ban on history-based price discrimination.²²

After the 2013 ban on history-based price discrimination, the interest rate difference between renewing and new consumers drops but remains statistically significant (Table 1.5). In economic terms, a renewing customer with an average mortgage of €150,000 pays between €70 (in 2014) and €167 (in 2015) more for the same type of loan than a new customer. Thus, the ban on history-based price discrimination significantly reduced the interest rate differences between renewing and new households, but did not eliminate them completely. The large increase in interest rates differences in 2015 is entirely due to one of the larger banks having differences between renewing and new customers comparable to pre-ban levels.²³ This finding is consistent with statements made by the AFM that there are some loopholes banks use to partially get around the ban.²⁴ Since, in particular for the years 2013 and 2014, the interest rate differences between renewing and new households are much smaller after than before the ban, I will assume that banks charge a single interest rate after the ban for simplicity.

²²See for example the newspaper article “Straf voor verlengen hypotheek moet van tafel”, *Het Financieele Dagblad*, 25 september 2010. This article reports pre-ban interest rate differences between 0.2 and 0.4 percentage points.

²³That is, removing this bank from the sample gives differences comparable to 2013 and 2014.

²⁴For example, many banks offer a discount if the mortgage contract is certified by a notary within a certain time frame. Since prolonging a mortgage does not require certification by a notary, such discounts are automatically not available to prolonging households (Autoriteit Financiële Markten 2015). Such discounts are now banned, but were still sometimes offered during my sample.

1.5 A model of mortgage demand and supply

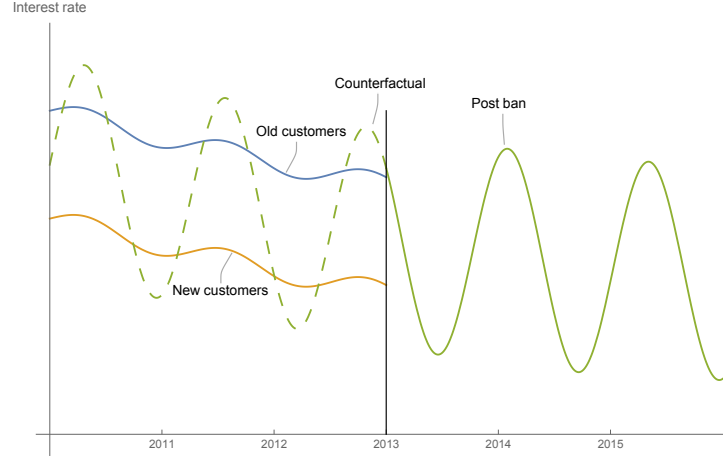
In this section, I develop a structural model of the Dutch mortgage market. I begin by describing the empirical strategy I use to identify the effects of history-based price discrimination. Then I describe the demand and supply side of the model. I conclude this section by introducing the new solution concept I use for the supply side, Sparse Markov Perfect Equilibrium.

1.5.1 Empirical strategy

To estimate the effect of the ban on history-based price discrimination, I use the empirical strategy depicted in Figure 1.2. I estimate a structural model of demand and supply. Since, as I explain in Section 1.4.2, I am better able to follow switching households over time after 2012 and switching behavior is a crucial element of my model, I estimate demand and supply on the data covering the *post-ban* period (2013-2015). Given estimates of banks' policy functions after the ban, I can predict what interest rates they would have counterfactually set *pre-ban* had history-based price discrimination already been banned then. Comparing these counterfactual interest rates with observed interest rates gives an estimate of the effect of the ban on interest rates. I then use the estimated demand and supply model to calculate the effect of the ban on consumer surplus and bank profits, respectively. This strategy estimates a causal effect to the extent that I am able to control for all relevant changes in the mortgage market around the time of the ban. Given that the market underwent many changes around the ban, this is a strong assumption. I describe the most important of these changes in Section 1.8.1. Here I also argue that these changes cumulatively did not have a large impact on mortgage pricing.

This approach differs from the more typical approach in empirical industrial organization where the estimated model is used to compute a counterfactual equilibrium. Because I observe both equilibria of interest (with and with history-based price discrimination), I can instead extrapolate banks' post-ban policy functions to the pre-ban period and compare with observed interest rates. This method has three advantages over computing a counterfactual. The first is that it is much simpler. Since the supply side of my model is a dynamic game, computing a counterfactual equilibrium is computationally challenging. Moreover, dynamic games typically have many equilibria. Thus, computing a counterfactual equilibrium raises the issue of equilibrium selection. Here, I can simply assume that banks keep playing the observed equilibrium. Finally, this method is arguably more credible since it does not require me to make assumptions on bank conduct when history-based price discrimination is legal.

Figure 1.2: Estimation strategy



Note: The model is estimated for the post-ban period. The reduced-form policy function estimates are then extrapolated to the pre-ban period. Comparing with observed pre-ban interest rates then gives the effect of the ban.

1.5.2 Demand

I estimate a standard discrete-choice model of demand.²⁵ Households choose which mortgage to purchase, I take the desired loan amount as given.²⁶ The relevant market in month t consists of households whose mortgages' fixed interest rate periods end, as well as households who move in that period and may purchase a mortgage if they buy rather than rent.²⁷ Denote all households that make a purchasing decision at time t by \mathcal{H}_t . A household chooses from up to J possible mortgages in the choice set \mathcal{C}_i . The choice set differs between households because, from 2013 onwards, a household can only purchase a non-amortizing mortgage if it previously had a non-amortizing mortgage (as I explain in Section 1.4.4).

Banks sell multiple mortgages. Denote by \mathcal{J}_b the set of mortgages sold by bank b . I

²⁵In practice, many households combine different loans, for example a savings mortgage with an annuity, into a single mortgage. I ignore this because the second mortgage almost always has the same repayment method (but has for example a different maturity) or is an annuity. Modeling the choice of the second (and potentially third) loan would impose significant difficulties, since it would require a model for the decision how much of the total loan sum to allocate to what type of loan in addition to a model of loan choice. However, there is little variation in the type of second loan so that the gains do not outweigh the benefits.

²⁶For renewing households, the desired loan amount *is* in fact given since they need to refinance their outstanding debt. Loan-to-value ratios for NHG loans are on average around 95%, which indicates that most first-time borrowers borrow close to as much as they can. (The maximum loan-to-value ratio is 102%). Therefore, the effect of the ban on loan amounts is not particularly interesting.

²⁷It is possible that a household ends its fixed interest rate period prematurely. This is costly, but it can sometimes be worthwhile if interest rates have dropped sufficiently since the start of the fixed interest rate period. However, ending a fixed interest rate period prematurely does not imply the need to switch: most households end up renewing the loan at their current bank. Therefore, I take the choice to end a fixed interest rate period as exogenous and model the mortgage choice of households that do so the same as of households whose fixed interest rate period ends contractually.

denote by $j = 0$ the outside option, which is sold by “bank” $b = 0$. The outside option includes not having a mortgage, as well as having a mortgage by an institution which is not in my sample.²⁸

The utility household i derives from mortgage j depends on the mortgage’s characteristics X_{jt} , its interest rate r_{jtn} , household income D_{it} and potentially switching costs s_{it} . I subscript the interest rate with $n \in \{0, 1\}$ to indicate the interest rates banks charge pre-ban to old and new customers, respectively. New customers include switching households and households that purchase their first mortgage. A mortgage is defined by the bank that sells it, the fixed interest rate duration (as in Table 1.2) and the payment method (as in Table 1.4). In other words, X_{it} contains bank, fixed interest rate duration, and payment method dummies.

Consumers face switching costs. A household incurs switching costs if it purchases a mortgage from a different bank—switching to a different type of mortgage by the same bank is free. This is a simplification since switching to a different type of mortgage will typically involve some costs. However, the main costs of switching to a different a bank—taxation, a notary and NHG insurance fees—do not need to be paid when staying at the same bank. Switching costs measure the monetary, hassle and time costs of switching. In addition, the switching cost parameter will pick up other frictions in the market as well, such as search costs and inattention. Disentangling between these sources of consumer inertia requires either variation in the magnitude of the different frictions or direct data on consumer search, both of which are unavailable in my setting.²⁹

The utility household i derives from purchasing mortgage j is

$$u_{ijt} = X_{jt}\Pi_{Dit} - (\alpha_{ijt}r_{jt1} + s_{it})\Delta_{ijt} - \alpha_{ijt}r_{jt0}(1 - \Delta_{ijt}) + \xi_{jt} + \varepsilon_{ijt}. \quad (1.1)$$

where Δ_{ijt} is a dummy denoting whether household i needs to pay switching costs to purchase mortgage j . Π_{Dit} and α_{ijt} are coefficients and ξ_{jt} and ε_{ijt} are error terms. The first term gives the utility the household derives from the mortgage characteristics X_{jt} . The second term contains the interest rate paid by new customers and switching costs s_{it} . Pre-ban, renewing customers pay a different interest and they never incur switching costs, which is reflected by the third term. The fourth term, ξ_{jt} , denotes the unobserved quality of mortgage j . As is usual in the empirical industrial organization literature, ξ_{jt} is allowed to be correlated with the interest rate. Since ξ_{jt} is unobserved, interest

²⁸In the remainder of this paper, I will often refer to households who have previously purchased a mortgage from some bank b . Such statements should always be understood to be valid for $b = 0$ as well, i.e. such statements also apply to households who previously purchased the outside option.

²⁹See Kiss (2017) for more on disentangling attention and switching costs and Honka (2014) on disentangling search and switching costs.

rates are endogenous, for which I will correct in the estimation of the demand model. The final term, ε_{ijt} , is an idiosyncratic error term which is assumed to be conditionally independent from all other variables in the utility specification.

The outside option gives utility

$$u_{i0t} = -s_{it}\Delta_{i0t} + \varepsilon_{i0t}.$$

In this specification, switching to and from the outside option is costly. The assumption that switching *from* the outside option is costly reflects that the same fees need to be paid when purchasing a new mortgage as when switching. Switching *to* the outside option is also costly as paying off a mortgage before its end date typically triggers a fine.

To make the demand model as flexible as possible, the model includes household-specific coefficients. The coefficients on mortgage characteristics, the coefficient on the interest rate and switching costs depend on household income D_{it} as follows:

$$\begin{aligned}\Pi_{Dit} &= \Pi_D D_{it}, \\ \alpha_{ijt} &= \exp\{\Pi_\alpha D_{it}\}, \\ s_{it} &= \exp\{\Pi_s D_{it}\}.\end{aligned}$$

The preceding formulation assumes that demand is static, despite the fact that households face an inherently dynamic problem: when fixing their interest rate, they have to form expectations about the interest rates they will face when they have to renew their interest rate. One reason I estimate a static demand model is precisely because I believe it is too complicated for most households to think ahead ten years and realize they might have to pay more then, let alone form correct expectations about the strategic behavior of banks in the future. In fact, survey evidence indicates that 60% of households do not even *consider* switching at the moment of renewal.³⁰ Therefore, it seems unreasonable to assume a significant fraction pays attention to dynamic considerations. A second reason for estimating a static demand function is that virtually all households in my sample make only a single decision during my sample period, so that any dynamic model of demand would be identified only by functional form.

³⁰The survey was conducted by the Dutch homeowners' association. Vereniging Eigen Huis. "Huseigenaren laten massaal geld liggen bij nieuwe rente hypotheek", January 3 2018. <https://www.eigenhuis.nl/actueel/nieuws/2018/01/03/09/00/huseigenaren-laten-massaal-geld-liggen-bij-nieuwe-rente-hypotheek>

1.5.3 Supply

Because the interest rate a bank sets today affects the number of captive consumers it has tomorrow, banks face a dynamic problem when setting their interest rates. I use a non-standard equilibrium concept, Sparse Markov Perfect Equilibrium, to solve the dynamic game between the banks. Before I introduce it below, I describe the constituent parts of the model: the state, how it evolves and banks' flow profit functions.

The state consists of two parts: cost shocks and previous sales. There is no private information: the complete state is known to all banks. At the beginning of every month, there is a shock to the common cost of funding i_t . In the Netherlands, mortgage funding comes from many different sources, including short-term and long-term deposits, money market funds and securitization. I use the interest rate on 10-year Dutch government bonds as a proxy for these costs, as this interest rate displays significant co-movement with observed mortgage interest rates (as can be seen in Figure 1.1).³¹

The second state variable consists of previous market shares, which banks need to take into account because households face switching costs. However, because of the heterogeneity in the demand function, banks not only need to take into account *the amount* of mortgages they sold in the past, but also *the type of consumers* they sold them to. In terms of the model, this is the case because the coefficients in households' utility Π_{Dit} and α_{it} , as well as switching costs s_{it} depend on household income D_{it} . Thus, there is in effect a different demand function for every type of household. Because of switching costs, the demand of a particular type of household depends on past market shares for this type. Therefore, total demand depends on past market shares of every type of household. In other words, the pay-off relevant state variable is the joint density of previous sales and household income that are in the market at time t . Denote this density by $f_t(b, D)$, where b is a random variable denoting which bank a household in \mathcal{H}_t purchased from previously and D is the distribution of household income. Denote the set of all state variables at time t by $\sigma_t = \{f_t(\cdot), i_t\}$.

Let r be the vector of all interest rates. The demand for mortgage j from consumers who previously bought from bank b

$$d_j(b, r) = M_t \int p_j(b, r, D) f(b, D) dD,$$

where p_j is the probability implied by the demand model that a household of type D whose current bank is b purchases mortgage j when interest rates equal r and M_t is the total demand for loans (in euros) at time t .

³¹I have also tried using the marginal rate on deposits as a proxy for the cost of funding. However, this is a worse proxy: when I include both the rate on Dutch government bonds and the marginal rate on deposits in the policy function regression in Section 1.6.1.2, the Lasso always selects only the interest rate of Dutch government bonds.

Denote by c_j the marginal cost of supplying a loan of €1. I let

$$c_{jt} = \gamma_{j0} + \gamma_{j1} + \gamma_{j2}i_t.$$

γ_{j0} measures any marginal cost of supplying loan j in addition to the cost of funding, for example the implied costs of pre-payment risk (which differ across loan types). Some loan types, such as savings or investment mortgages, are commonly sold together with other high-margin products such as life insurance. γ_{j0} also measures this implicit cross-subsidy. I restrict γ_{j0} to be the same for all loans with the same payment method sold by the same bank, e.g. all annuities sold by bank b have the same γ_{j0} . γ_{j1} and γ_{j2} measure the cost of funding. Similarly, I restrict γ_{j1} and γ_{j2} to be the same for all loans with the same fixed interest rate duration sold by the same bank, e.g. all 10-year loans sold by bank b have the same γ_{j1} and γ_{j2} . Thus γ_{j1} and γ_{j2} measure the differences in cost of funding of loans with different fixed interest periods as a function of the average cost of funding.

Denote by r_b and r_{-b} the interest rates set by bank b and its competitors, respectively. The flow profits of bank b in state σ are

$$\pi_b(\sigma, r) = \sum_{j \in \mathcal{J}_b} \left\{ (r_{j0} - c_j) d_j(b, r_b, r_{-b}) + \sum_{b' \neq b} (r_{j1} - c_j) d_j(b', r_b, r_{-b}) \right\}.$$

The first term contains the profits from customers who do not switch, the second term from customers who do.

Given the state and banks' interest rates, the future state can be calculated as follows. At time t , the proportion of households with mortgage j whose mortgage will expire in τ months is equal to the weighed sum of the proportion at time $t - 1$ of households with that mortgage expiring in $\tau + 1$ months and the market share at time $t - 1$ of mortgage j among mortgages with a duration of τ months. The weights are the number of mortgages expiring in $\tau + 1$ months and the number of expired mortgages at time $t - 1$, respectively. The evolution of the joint density of previous purchases and household characteristics, evaluated at a point (j, D) , can be written as

$$f_{t+\tau}(b, D) = \frac{|\mathcal{H}_{t+\tau}|}{|\mathcal{H}_{t+\tau}| + |\mathcal{H}_t|} f_{t+\tau}(b, D) + \frac{|\mathcal{H}_t|}{|\mathcal{H}_{t+\tau}| + |\mathcal{H}_t|} \sum_{j \in \mathcal{J}_b} \sum_{b'} \phi_j(\tau) p_j(b', r, D) f_t(b, D),$$

where $\phi_j(\tau)$ is an indicator function that equals 1 if and only if product j has a fixed interest duration of τ months. $|\mathcal{H}|$ denotes the number of elements in \mathcal{H} , i.e. $|\mathcal{H}_t|$ is the number of households that make a purchasing decision at time t . Combining this transition function for all points (j, D) and all $\tau = 1, \dots$, gives the full transition function for f :

$$f_{t+\tau} = \Gamma_t(s, r).$$

1.5.4 Sparse Markov Perfect Equilibrium

The typical solution concept for dynamic games is Markov Perfect Equilibrium (Maskin and Tirole 2001, MPE hereafter). An MPE is a sub-game perfect equilibrium in which agents' strategies are constrained to be functions of only the payoff-relevant state.³² That is, an MPE consists of policy functions $\rho_b(\sigma)$ such that

$$\rho_b(\sigma) = \arg \max_{\rho} \pi_b(\sigma, r, \rho_{-b}(\sigma)) + \beta \mathbb{E}[V_b(\Gamma(\sigma, r))]$$

for all possible states σ . Here,

$$V_b(\sigma) = \pi_b(\sigma, \rho_b(\sigma), \rho_{-b}(\sigma)) + \beta \mathbb{E}[V_b(\Gamma(\sigma, r))]$$

is bank b 's Bellman equation and $\Gamma(\sigma, r)$ the evolution law.

In the supply model derived above, the payoff-relevant state is infinite-dimensional: banks have to keep track of the joint density of previous market shares and household demographics. This creates a challenge when estimating the model as existing methods (e.g. Bajari, Benkard, and Levin (2007)) require a first stage in which policy functions are regressed on state variables. When the number of state variables is larger than the number of observations of the policy function, such a regression is impossible. Since I have an infinite number of state variables, I cannot proceed as is standard.

In general, the pay-off relevant state will be large when the model, or parts of it, allows for rich heterogeneity. Typically, this kind of rich heterogeneity is required to match the patterns found in the data. The most common example of this is in demand estimation, where it is well known that the simple logit model cannot capture typical demand elasticities.³³ Normally, the researcher then faces a trade-off: richer heterogeneity in one part of the model (for example in the demand function) increases the size of the state space in another (for example in the supply side). Here, I could reformulate the model in such a way that the state space is finite-dimensional, for example by having only two types of households, rich and poor, in the demand function. However, this would reduce the flexibility of the demand function a lot.³⁴ Because it is well-known that the shape of the demand function is a crucial determinant of the welfare effects of third-degree price discrimination (Holmes 1989), it is important that I allow for sufficient flexibility of the demand function. By introducing techniques to

³²That is, agents' strategies cannot be functions of play in previous' periods, except insofar that play changes the current period's state.

³³See, for example, Akerberg et al. (2007) for a discussion of this issue.

³⁴Moreover, to obtain flexible estimates of policy functions, it is typically required to include functions of state variables, such as higher-order polynomials or interactins. Given the number of observations I have, this would already be impossible when the heterogeneity is restricted to two groups.

deal with large state spaces in the estimation, I hope to relieve this tension: it is possible to have rich heterogeneity in one part of the model without creating insurmountable challenges when estimating the part of the model with the dynamic game.

To solve the challenge of a large state space, I combine techniques from machine learning with micro-economic theory. The theoretical part consists of a new solution concept that allows easier estimation of games with large state spaces: Sparse Markov Perfect Equilibrium (SMPE). As I explain in the section on estimation below, SMPE is particularly amenable to estimation using machine learning techniques. The main difference with a standard Markov Perfect Equilibrium is that I relax the assumption that information on the state is free. In an SMPE, agents then optimally pay attention to a subset of the state. As a result, the domains of their policy functions have a smaller dimension, making estimation and calculation much simpler. This concept is not only *computationally* attractive, but also *behaviorally*. Banks do not just know the state they are in, they need to perform some kind of market research. Such research is costly. Therefore, they will only try to figure out those state variables of which knowledge has a sufficiently large impact on their profits. For example, the MPE of the supply model implies that banks' policies are a function not just of past market shares, but also of market shares across households with an income of €30,000, of €31,000, and so on, as the full density of household demographics and past market shares is payoff-relevant. Most likely, it does not pay off for banks to invest in such detailed knowledge.³⁵ SMPE formalizes this intuition.

1.5.4.1 Sparse maximization

Before I introduce SMPE, I give a brief introduction to sparse maximization. Sparse maximization, introduced for a single decision maker by Gabaix (2014, for static settings) and Gabaix (2017, for dynamic settings), is a simple method to model inattention. The basic idea is that a decision maker pays attention to a variable if the benefits of paying attention exceed the costs. The sparse maximization operator then defines the benefits of paying attention in such a way that the model becomes tractable.

Sparse maximization works as follows. For notational simplicity, consider a single bank.³⁶ Denote by σ the vector of pay-off relevant state variables. The bank wants to maximize $v(\sigma, r)$, where r is the vector of interest rates the bank sets. In a static context, $v = \pi(\sigma, r)$, in a dynamic context $v = \pi(\sigma, r) + \beta \mathbb{E}[V(\Gamma(\sigma, r))]$, where $V(\cdot)$

³⁵Moreover, the manager in charge of setting interest rates may be cognitively limited and unable to process the full information, instead choosing to focus on the most important state variables. As shown by Gabaix (2014), such bounded rationality also leads to sparse maximization.

³⁶I cover the extension to games below.

is the continuation value. Let $m_i \in \{0, 1\}$ indicate whether the bank pays attention to state variable σ_i .³⁷ There is a fixed cost of paying attention of $\kappa \geq 0$ per state variable.³⁸ In Gabaix (2014, 2017), this cost is interpreted as a psychological cost. In my setting, however, a more natural interpretation is the real cost of obtaining or processing information on a state variable, for example due to cost of market research or revenue management systems.

Every period, the bank forms a *sparse state*, $\hat{\sigma}$. It knows the true value of every state variable it pays attention to. For variables it does not pay attention to, it assumes they equal some default value σ^d . (This default is specified by the researcher. Below I specify how I define the default state in this application.) Thus, the bank's sparse state is

$$\hat{\sigma}_i = \begin{cases} \sigma_i & \text{if } m_i = 1, \\ \sigma_i^d & \text{if } m_i = 0 \end{cases}$$

for every state variable σ_i . The bank chooses its interest rates to maximize its profits in the sparse states:

$$\rho(\hat{\sigma}) = \arg \max_r v(\hat{\sigma}, r) = \arg \max_r v(m\sigma + (1 - m)\sigma^d, r).$$

How to choose the attention vector m ? For a given attention vector m , the bank expects to lose

$$\mathbb{E}[v(\sigma, \rho(\hat{\sigma})) - v(\sigma, \rho(\sigma))] \quad (1.2)$$

from not paying attention. Here, $\rho(\sigma) = \arg \max_r v(\sigma, r)$ is the vector of optimal interest rates in the true state σ . Therefore, the bank should choose its attention vector m to minimize

$$\mathbb{E}[v(\rho(\sigma, \hat{\sigma})) - v(\sigma, \rho(\sigma))] + \kappa \sum_i m_i.$$

However, this problem is more complicated than maximizing $v(\cdot)$ under full attention. Therefore, it does not offer a simplification. The idea of sparse maximization is to replace this (very intractable) problem by a (more tractable) second order Taylor expansion around the default state σ^d . In the default state, the bank's profit maximizing interest

³⁷Gabaix (2014, 2017) also considers values of m_i between 0 and 1. In this case, the bank would display partial attention. I do not consider this case, since it is not required to make the state space simpler. Moreover, partial inattention cannot directly be identified from the data using the methods I introduce in Section 1.6.1. However, once banks are restricted to display either full or no inattention, identification becomes particularly simple. This assumption is equivalent to a functional form assumption on the cost of attention, see also Footnote 38.

³⁸Gabaix (2014, 2017) also considers other cost functions. Many of these cost functions lead to partial inattention. Since the focus of this paper is not on the cost function (for example, I do not estimate the cost of attention), I pick the simplest cost function that leads to sparsity.

rates are

$$r^d = \rho(\sigma^d) = \arg \max_r v(\sigma^d, r).$$

Gabaix (2014, 2017) calls r^d the “default action”. The second order Taylor expansion of (1.2) with respect to m_i is³⁹

$$\Lambda_i \equiv -\frac{1}{2} \mathbb{E}[(\sigma_i - \sigma_i^d)^2] r_{\sigma_i}^T v_{rr} r_{\sigma_i}, \quad (1.3)$$

with

$$r_{\sigma_i} \equiv \left. \frac{\partial \rho(\hat{\sigma})}{\partial \hat{\sigma}_i} \right|_{\hat{\sigma}=\sigma^d},$$

$$v_{rr} \equiv \left. \frac{\partial^2 v(\sigma^d, r)}{\partial r^2} \right|_{r=r^d}.$$

The bank pays attention to σ_i if the cost of inattention Λ_i is greater than the cost of attention κ . The expression for Λ_i shows that the cost of inattention is larger when

1. the state variable σ_i shows more variation, or,
2. knowing the true value of σ_i has a greater impact on the optimal interest rates set by the bank.

The sparse max operator is then defined as follows.⁴⁰

Definition 1 (Sparse max operator (Gabaix 2017)). *The sparse max, $\text{smax}_r v(r, \sigma)$, is defined by the following procedure.*

1. Choose the attention vector m^* :

$$m^* = \arg \min_m \sum_i (1 - m_i)^2 \Lambda_i + \kappa \sum_i m_i,$$

Given m^* , a bank forms a sparse state $\hat{\sigma} = m^* \sigma + (1 - m^*) \sigma^d$.

2. Choose the interest rates

$$r = \arg \max_r v(\hat{\sigma}, r).$$

³⁹The first term of the expansion equals zero because it is equivalent to the first order condition of the bank. The expectation follows from rewriting $\hat{\sigma}_i = m_i \sigma_i + (1 - m_i) \sigma_i^d$.

⁴⁰To be precise, this is what (Gabaix 2017) calls the “sparse max operator without budget constraint”. I do not require the version with budget constraint.

1.5.4.2 SMPE: Definition

I now extend sparse maximization to dynamic games. An SMPE is an MPE, except using sparse maximization. In every state, every bank maximizes its profits as if it were in the corresponding sparse state, given its beliefs on the interest rates its competitors set.

I define the default state σ^d as follows. I assume that banks substitute the long-run average for state variables they do not pay attention to.⁴¹ Therefore, bank b 's sparse state can be defined as follows.

Definition 2 (Sparse state). *Bank b 's sparse state σ_b is*

$$\hat{\sigma}_{bi}(\sigma_i) = \begin{cases} \sigma_i & \text{if } m_{b,i} = 1, \\ \mathbb{E}[\sigma_i] & \text{if } m_{b,i} = 0, \end{cases}$$

for all state variables i , where $\mathbb{E}[\sigma_i]$ is the long-run average of state variable i and m_b is b 's attention vector.

Since different banks can pay attention to different variables and form different sparse states, the question arises what banks' beliefs should be on their competitors' actions. A natural assumption is that when a bank behaves as if the true state equals its sparse state, it assumes that its competitors do so as well. Moreover, beliefs have to be consistent in equilibrium.

Assumption 1 (Consistency of beliefs). *A bank's beliefs on its competitors actions are consistent with equilibrium and with the structure of its own sparse state. Bank b believes its competitors behave as if the true state is equal to b 's sparse state:*

$$\hat{\rho}_{-b}(\sigma) = \rho_{-b}(\hat{\sigma}_b(\sigma)),$$

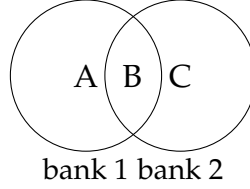
where $\hat{\rho}_{-b}(\cdot)$ is bank b 's belief on its competitors actions and $\rho_{-b}(\cdot)$ denotes b 's competitors' equilibrium interest rates.

As an example of belief formation, consider the situation in Figure 1.3. Bank 1 pays attention to state variables A and B (and not C). Therefore it assumes that the state is equal to $(A, B, \mathbb{E}[C])$, i.e. it substitutes the long-run average of C for its true value. By Assumption 1, bank 1 believes that bank 2 behaves as if the state is $(A, B, \mathbb{E}[C])$ as well. To be consistent with equilibrium, Assumption 1 requires that bank 1 believes bank 2's action in (A, B, C) is equal to its equilibrium action when the *true* state is $(A, B, \mathbb{E}[C])$.⁴²

⁴¹I rescale state variables that are past market shares so that the sum of market shares equals one.

⁴²Note that bank 2's sparse state does not include A and that therefore, the requirement that A is at

Figure 1.3: Illustration of belief formation in SMPE



Note: In this example, there are three variables in the true state—A, B and C. Bank 1 pays attention to A and B, bank 2 to B and C. Therefore, changes in C do not change bank 1’s belief on bank 2’s interest rates. Similarly, changes in A do not change bank 2’s belief on bank 1’s interest rates.

To define an equilibrium, I require that for every possible state σ , banks’ policy functions are sparse maximizers of their discounted expected profits. Thus, in equilibrium bank b ’s Bellman equation is

$$V_b(\hat{\sigma}_b) = \text{smax}_r \pi_b(r, \rho_{-b}(\hat{\sigma}_b), \hat{\sigma}_b) + \beta \mathbb{E} [V_b(\Gamma(\hat{\sigma}_b, r, \rho_{-b}(\hat{\sigma}_b)))], \quad (1.4)$$

where ρ_{-b} are the policy functions of b ’s competitors and β is the discount rate. Let $\hat{\Sigma}_b$ be the set of bank b ’s possible sparse states. An SPME is then defined as follows.

Definition 3 (Sparse Markov Perfect Equilibrium). *A Sparse Markov Perfect Equilibrium consists of policy functions $\rho_b : \hat{\Sigma}_b \rightarrow \mathbb{R}^{|\mathcal{J}_b|}$, such that*

$$\rho_b(\hat{\sigma}_b) = \text{smax}_r \pi_b(\hat{\sigma}_b, r, \rho_{-b}(\hat{\sigma}_b)) + \beta \mathbb{E} [V_b(\Gamma(\hat{\sigma}_b, r, \rho_{-b}(\hat{\sigma}_b)))],$$

for all banks $b \in \mathcal{B}$ and all states $\hat{\sigma}_b \in \hat{\Sigma}_b$, where $V_b(\cdot)$ is given by (1.4).

To formally investigate SMPE’s theoretical properties is outside the scope of this paper, where the focus is on the empirical application.⁴³ However, it seems likely that the usual multiplicity problem of MPE’s is even worse for SMPE’s: not only can there be many equilibria for given attention vectors, but it is also possible that there are equilibria which differ with respect to the state variables banks pay attention to. Since I assume that banks keep playing the estimated equilibrium, this is not an issue in this application. A second question is whether SMPE can be thought of as an approximation to MPE. There is some reason to believe this might be the case: Gabaix (2017) shows that for finite-horizon single agent problems, the policy and value functions under sparse

the true value is superfluous: holding B and C constant, changes in A do not change bank 2’s action. Thus, a state variable changes the belief of one bank on another bank’s action only if that variable is in both banks’ sparse states. In this case, since B is the only state variable that is in both banks’ sparse states, only changes in B change banks’ beliefs. However, changes in A do change bank 1’s behavior (since A is in its sparse state) and changes in C change bank 2’s behavior.

⁴³One hurdle in deriving the theoretical properties is that banks do not have rational expectations. As stressed by Gabaix (2017), this creates methodological challenges because backwards induction cannot be used.

maximization differ only in second terms from the policy and value functions under rationality. I have not investigated whether this result extends to infinite-horizon games, but this single player finite-horizon result at least gives some indication that SMPE might be an approximation to MPE.

1.6 Identification and estimation

In this section, I discuss the identification and estimation of my structural model. I start with the main methodological contribution of this paper: the identification and estimation of inattention. Then, I discuss the estimation of the demand and supply side of my model.

1.6.1 Sparse states and policy functions

1.6.1.1 Identification of sparse states

The intuition behind the identification of sparse states is simple: a bank pays attention to a state variable if and only if its policy function is a function of that state variable.⁴⁴ This is true since the sparse max operator picks precisely those state variables to pay attention to that have the largest impact on a bank's optimal action. Conversely, if one observes that a bank's interest rates do not vary with a certain state variable, it must not have paid attention to that variable: there is no point in a bank paying attention to a state variable if it knowing the value does not have an impact on its interest rates. Indeed, the cost of inattention Λ_i in (1.3) is increasing in r_{σ_i} , the derivative of a bank's optimal interest rates with respect to the state, so that a bank pays attention to a state variable only when knowing its true value sufficiently changes its actions.

Another way to look at this is by observing that the sparse max operator in effect truncates the policy functions. Gabaix (2014) shows that if policy functions are linear when agents pay full attention, with the coefficient on σ_i equal to a_i , the coefficient on σ_i under sparse maximization is simply⁴⁵

$$a_i \mathcal{I} \left(a_i^2 \geq 2 \frac{\kappa^2}{\text{Var}(\sigma_i)} \right).$$

Here, $\mathcal{I}(\cdot)$ is the indicator function. Therefore, a bank pays attention to a state variable σ_i if the policy function when attention is free varies "enough" with σ_i .

⁴⁴For legibility, I write paying attention to a state variable. It should be understood that by this I also mean any function of state variables, such as particular moments of the previous market share distribution f .

⁴⁵The same holds for the first-order Taylor approximation.

Since policy functions are observed by the econometrician, all that is required to identify sparsity is to select those state variables which have the largest effect on a bank's policy function. This is precisely what variable selection methods from the machine learning literature do: estimating a banks' policy function with a suitable penalized regression method is sufficient to identify its sparse state.

1.6.1.2 Estimation of sparse states and policy functions

To estimate which variables are in a bank's sparse state, I employ the Lasso. The Lasso augments the standard ordinary least squares (OLS) objective function with a penalty term on the coefficients. The result of this penalty term is that many coefficients will be set at exactly zero. Thus, the Lasso gives a sparse regression, where only the most important variables will have non-zero coefficients associated with them.

Banks sell multiple types of mortgages, each having a different interest rate. An SMPE implies that the policy functions of all the mortgages of the same bank should depend on the same sparse state. This is because paying attention has a fixed cost per state variable per firm: thus, it never pays for a bank to ignore a state variable for some mortgages but not for others. In other words, it is required to select the same variables for all policy functions of the same bank. To accomplish this, I use what the machine-learning literature calls a multi-task Lasso (Zhang 2006; Liu, Ji, and Ye 2009; Obozinski, Taskar, and Jordan 2010). The multi-task Lasso simultaneously estimates the coefficients of multiple models, imposing the same sparsity structure on all of them.⁴⁶

The multi-task Lasso works as follows. Group bank b 's interest rates in a $T \times |\mathcal{J}_b|$ matrix R_b , where the (t, j) -element of R_b is r_{jt} , the interest rate charged for mortgage j .⁴⁷ Let X be a $T \times K$ matrix of possible regressors, where K can be larger than the number of observations T . The multi-task Lasso is defined as

$$\hat{B} = \arg \min_B \frac{1}{2T} \|R - XB\|_{\text{Fro}}^2 + \lambda \|B\|_{21},$$

where $\|A\|_{\text{Fro}} = \sqrt{\sum_{ij} a_{ij}^2}$ is the Frobenius norm and $\|A\|_{21} = \sum_i \sqrt{\sum_j a_{ij}^2}$ is the $\ell_1 \ell_2$ norm. $\lambda > 0$ is the regularization parameter that generates sparsity. The larger λ , the sparser the set of selected variables. When $\lambda = 0$, the multi-task Lasso reduces to a separate OLS regression for each policy function. The j 'th column of the estimated

⁴⁶An additional advantage of using the multi-task Lasso is that it increases the number of effective observations in the Lasso regression. This can matter since I only have 36 observations (three years of monthly data) per mortgage. However, every bank sells 24 different mortgages, so that when I use a multi-task approach I have $24 \times 36 = 864$ observations per bank to identify the sparsity of its policy functions.

⁴⁷I only estimate policy functions on the post-ban data, so there is no need to distinguish between the interest rates charged to new and renewing customers.

coefficients \hat{B} , call it \hat{B}_j , contains the estimated parameters for r_j . The multi-task Lasso constrains these estimates such that all \hat{B}_j 's of the same bank have the same sparsity structure. In other words, \hat{B} will contain a zero row for every covariate that is not included.⁴⁸

The most common method to select the penalty parameter λ is k -fold cross-validation.⁴⁹ This procedure however assumes that the data are i.i.d. and banks' interest rates display serial correlation (see Figure 1.1). Therefore, I use time series cross-validation: I use to the first t observations to estimate the model and the $t + 1$ 'th observation to calculate the out-sample-fit. I do this for every $t = 1, \dots, T - 1$ and take the value of λ that gives the best average out-of-sample fit over all values of t .

After using the Lasso to estimate the sparsity of banks' policy functions, I use OLS to estimate their coefficients as in Belloni and Chernozhukov (2013). Using post-Lasso OLS is important, since the Lasso alone shrinks coefficients towards zero too much, leading to implausible counterfactual interest rates.⁵⁰

In X I include (functions of) the payoff-relevant state variables i_t and $f_t(b, D)$. The Lasso only works well if all independent variables are on the same scale. Therefore, I standardize them. One problem is that $f_t(b, D)$ is a density and thus infinite-dimensional. Therefore, I discretize this state variable by including as variables the market shares of households within certain brackets as follows. First, I include the overall market shares, $\int f_t(b, D)dD$. Then, I split the sample into two groups (rich and poor), and add the conditional market shares $\int_{d > E[D]} f_t(b, D)dD$ and $\int_{d \leq E[D]} f_t(b, D)dD$. Then, I divide the sample in four groups, etcetera. These variables are however highly correlated. When the independent variables are highly correlated, the Lasso can become unstable and pick a somewhat arbitrary variable from the set of correlated variables. For example, it might pick the market share of bank A among the poorest quarter of the population but not the overall market share of bank A. This is largely inconsequential for the fit, but makes the interpretation of the estimates somewhat difficult: it does not

⁴⁸One downside of this approach is that the multi-task Lasso also chooses the same functional form for every policy function when higher-order polynomials and interactions of variables are included. I have experimented with running a separate Lasso for every policy function to estimate its functional form after using the multi-task Lasso to pick which variables are in its domain, but this did not lead to significantly different results.

⁴⁹Cross-validation is a data-driven approach to choosing λ . The data set is split into a training and test set. For different values of λ , the model is estimated on the training set. The estimated model that gives the best out-of-sample fit on the test set is chosen. In k -fold cross validation, this is done k times and the model with the best average out-of-sample prediction over the k folds is chosen.

⁵⁰I have also tried using the fitness-based threshold Lasso combined with post-Lasso OLS, which according to Belloni and Chernozhukov (2013) is superior to just post-Lasso OLS. However, the thresholding did not induce any additional sparsity in my case so that the fitness-based threshold Lasso is equivalent to ordinary Lasso.

make a lot of sense for a bank to pay attention to the market share amongst a subsection of the population but not the overall market share.

Therefore, I employ the following procedure. First, I include the marginal market shares of deciding households, $\int f_t(b, D)dD$, as variables. I run the multitask post-Lasso OLS as described above using a second-degree polynomial of these variables. Then, I split the sample into two groups (rich and poor), and add the conditional market shares $\int_{d>E[D]} f_t(b, D)dD$ and $\int_{d\leq E[D]} f_t(b, D)dD$. In the spirit of the Frisch-Waugh-Lovell theorem, I run a multitask Lasso regression of the residuals of the first regression on the residuals of a regression of the split market shares on the non-split market shares to see whether they offer additional explaining power. If they do, I split every group again—so you get the market shares for the four income quantiles—and use the same procedure until no variables are added.

1.6.2 Demand

I estimate the demand model using maximum likelihood. To control for the endogeneity of interest rates I use the control function approach (Petrin and Train 2010). I use this approach instead of the more typical approach of Berry, Levinsohn, and Pakes (1995, BLP hereafter) for the following reasons. First, I want to exploit that I have so-called “micro-level” data, instead of the “market-level” data that BLP requires. BLP can be augmented to incorporate micro-level data as well (Berry, Levinsohn, and Pakes 2004), but the control function approach is more efficient since it is based on maximum likelihood rather than the generalized method of moments like BLP. Second, many mortgages are not sold every month. Thus, would I use BLP, I would face the “zero market share” problem.⁵¹ However, zero market shares are not problematic for the control function approach.

To apply the control function method, I replace the unobserved product quality ξ_{jt} in consumers’ utility specification (1.1) by so-called control functions, $\hat{\mu}_{jt}$:

$$\tilde{u}_{ijt} = X_{jt}\Pi_D D_{it} - (\alpha_{it}r_{jt1} + s_{it})\Delta_{ijt} - \alpha_{it}r_{jt0}(1 - \Delta_{ijt}) + \psi\hat{\mu}_{jt} + \varepsilon_{ijt}.$$

Here ψ is an additional parameter to be estimated. The control functions are residuals from a first stage regression. I use the policy function regressions from the previous section as the first stage. This means I use the state variables of my model, banks’ common cost of funding and previous market shares, as instruments. The cost of funding is a valid instrument since it shifts banks’ supply curves but does not directly

⁵¹The problem is that BLP is based on inverting a system of equations based on the logarithms of the market shares. Since the logarithm of zero is undefined, BLP does not work when some products have zero market shares.

affect demand. Previous market shares are a valid instrument if they are uncorrelated with current unobserved product qualities. Since market shares from previous periods are obviously correlated with unobserved product qualities from that same period, this requires that there is not too much autocorrelation in unobserved product qualities. Most renewers purchased their mortgage ten years prior. Therefore, there should be no autocorrelation at a lag of ten years.⁵²

These instruments identify the effect of interest rates on demand. The identification of the other parameters of the demand model is standard. Differences in market shares between mortgages identify the mean contribution to utility of every product characteristic. Differences in purchases between households with different incomes identify how the coefficients in the demand vary with household income.

To estimate the model, I make the usual assumption that the idiosyncratic utility shocks ε_{ijt} follow a Type I Extreme Value distribution, so that the probability that household i purchases mortgage j can be written as

$$p_{ijt} = \frac{\exp\{\tilde{u}_{ijt}\}}{\sum_{k \in \mathcal{C}_i} \exp\{\tilde{u}_{ikt}\}}.$$

As explained in Section 1.4, I have a choice-based sample: households that purchase the outside option have a smaller probability of being in my sample than households that purchase an inside good. Manski and Lerman (1977) show that reweighing the likelihood function gives consistent estimates in this case. Let q_i be the sampling probability of household i . The log-likelihood is then

$$\log \mathcal{L}(\Pi_D, \Pi_{ac}, \Sigma) = \sum_t \sum_{i \in \mathcal{H}_t} \frac{\log p_{ijt}}{q_i}.$$

When calculating standard errors, I account for first-stage estimation error using the methods of Karaca-Mandic and Train (2003).

1.6.3 Supply: marginal cost estimation

I estimate the supply side to get banks' marginal costs. The estimation of the supply side game consists in two stages. In the first stage, I estimate policy functions and which variables banks pay attention to—this is explained in Section 1.6.1. I now show how to use the first stage policy function estimates to estimate banks' marginal costs.

To estimate marginal costs, I use banks' stochastic Euler equations. It turns out that banks' number of controls (interest rates) is larger than the dimension of their sparse

⁵²There is no issue if there is autocorrelation between one year's quality and the next, as there will undoubtedly be.

states. As a result, I can generate “moments” using only data from the current period. To see why, consider an example of a single agent that has two controls r_1, r_2 and one (perceived) state variable σ . Its Bellman equation is

$$V(\sigma) = \max_{r_1, r_2} \pi(\sigma, r_1, r_2) + \beta \mathbb{E} V(\Gamma(\sigma, r_1, r_2)).$$

As is typical for dynamic models, it is possible to reformulate the Bellman equation so that the bank directly chooses the future state σ' :⁵³

$$V(\sigma) = \max_{\sigma'} \tilde{\pi}(\sigma, \sigma') + \beta \mathbb{E} V(\sigma'), \quad (1.5)$$

$$\tilde{\pi}(\sigma, \sigma') = \max_{r_1, r_2 \text{ s.t. } \Gamma(\sigma, r_1, r_2) = \sigma'} \pi(\sigma, r_1, r_2). \quad (1.6)$$

The first order conditions of the maximization problem (1.6) are

$$\frac{\partial \pi}{\partial r_i} - \mu^T \frac{\partial \Gamma}{\partial r_i} = 0 \quad (1.7)$$

for $i = 1, 2$, where μ is the Lagrange multiplier. Except for μ and any unknown parameters of the model, all quantities in this expression are observed or can be calculated. Since there are two first order conditions and one multiplier, one parameter of the model can be estimated without using the first order conditions implied by the Bellman equation (1.5).

The intuition for this result is as follows. Say that $\frac{\partial \Gamma}{\partial r_1} = \frac{\partial \Gamma}{\partial r_2}$. If $\frac{\partial \pi}{\partial r_1} > \frac{\partial \pi}{\partial r_2}$, the bank has the following profitable deviation. It can increase r_1 and simultaneously decrease r_2 by ε . Because $\frac{\partial \Gamma}{\partial r_1} = \frac{\partial \Gamma}{\partial r_2}$, this leaves next period's state unaltered. However, since $\frac{\partial \pi}{\partial r_1} > \frac{\partial \pi}{\partial r_2}$, this deviation strictly increases today's profits. Therefore, if $\frac{\partial \Gamma}{\partial r_1} = \frac{\partial \Gamma}{\partial r_2}$, it is required that $\frac{\partial \pi}{\partial r_1} = \frac{\partial \pi}{\partial r_2}$. Generalizing this logic, optimality requires that in every month and for every interest rate a bank sets, the derivative of its profit function with respect to that interest rate is proportional to the derivative of the evolution law.

The same result can be derived for the full model. Let S_b be the number of variables in bank b 's sparse state $\hat{\sigma}_b$. The equivalent of (1.7) is

$$g_{bt}(\gamma, \lambda) \equiv \frac{\partial \pi_b}{\partial r_b} - \frac{\partial \Gamma_b}{\partial r_b} \mu_{bt} = 0, \quad (1.8)$$

where μ_{bt} contains the S_b multipliers of bank b in month t . Post-ban, there are $|\mathcal{J}_b|$ first order conditions: one for every interest rate. If $S_b < |\mathcal{J}_b|$, as is the case, these first order conditions contain additional information that can be used to estimate banks' marginal

⁵³The equivalence becomes immediately obvious by taking the first order conditions of both formulations.

costs.⁵⁴ I derive the functional forms of these first order conditions in Appendix 1.B.2. It is important to note that, following the definition of an SMPE, a bank maximizes its profits conditional on the policy functions of its competitors in its own sparse state. Thus, when calculating the first order conditions of bank b , one must use the estimated policy functions of its competitors evaluated in b 's sparse state.

Using only the first order conditions implied by (1.6) and not those implied by (1.5) has various benefits. First, it does not require the direct calculation of expectations, nor their estimation by substitution of future observed values. This is especially important in my application as in an SMPE banks have non-rational expectations and using this method allows estimation without specifying how banks *do* form expectations.⁵⁵ Also, it is possible to estimate marginal costs without specifying banks' discount factor β , which is typically not identified without further exclusion restrictions. Of course, these benefits come at the cost of a loss of efficiency if the model is specified correctly and the discount factor is identified or known.

I estimate marginal costs as follows. For every candidate parameter vector γ_b and period t , I find the Lagrange multipliers $\mu_{bt}(\gamma_b)$ —subscripted to indicate their dependence on the trial parameters—that solve bank b 's first order conditions (1.8). Since this is an overdetermined system of equations, it will not be possible to find μ_{bt} that solve (1.8) for all (or any) γ_b . Therefore, I find $\mu_{bt}(\gamma_b)$ by OLS. I then search over parameters γ_b to minimize the average (over months) first order conditions, i.e. I solve the minimum distance objective

$$\min_{\gamma_b} \left(\frac{1}{T} \sum_{t=1}^T g_{bt}(\gamma_b, \mu_{bt}(\gamma_b)) \right)^T \left(\frac{1}{T} \sum_{t=1}^T g_{bt}(\gamma_b, \mu_{bt}(\gamma_b)) \right).$$

The value of γ_b that minimizes this objective function gives estimates of bank b 's marginal cost parameters.

⁵⁴I stress that there is nothing in the definition of an SMPE that requires this to be the case; it is what I find empirically. For example, one could also estimate the SMPE of a model with one control, so that there are never any additional degrees of freedom whatever the sparsity structure of banks' policy functions. When this method is not applicable, value function iteration as in Bajari, Benkard, and Levin (2007) can be used instead.

⁵⁵In the definition of an SMPE, I have specified that banks form expectations according to the value function induced by (1.4). This means that banks are aware that they are sparse maximizers. However, as (Gabaix 2017) discusses, it is also possible that banks are "naive" and expect to obtain the rational value function (without sparse maximization). Estimating marginal costs in this way is thus robust to this assumption.

1.7 Results

This section presents my results. I start with the estimates of the structural model: policy functions, demand functions and marginal costs. Then, I answer the main question of this paper: what is the effect of the ban on history-ban price discrimination on consumer surplus, bank profits and welfare?

1.7.1 Model primitives

1.7.1.1 Sparse states and policy functions

Table 1.6 contains the estimated structure of banks' policy functions, and therefore also the variables in their sparse states. The multitask Lasso leads to a significant reduction of the size of the state space: the actual state is infinite-dimensional, the sparse states contain at most five variables. Yet, the fit is excellent. The R^2 's imply that these few variables are able to explain most variance in observed interest rates. Remember that the residuals from these policy function estimates are used as control functions in the demand estimation. Therefore, the high R^2 's moreover imply that the model's state variables are highly relevant instruments.

Every bank's policy function depends on the cost of funding, as expected. In addition, their policy functions depend on the overall market shares of one to three other banks. No bank's policy function depends on market shares conditional on households falling into a certain income bracket. Although I cannot report the results on which this is based because of data confidentiality, banks are more likely to pay attention to the largest bank in the market or to price-fighting banks that tend to offer low interest rates.

1.7.1.2 Demand

Table 1.7 contains estimates of the demand model. The estimates imply that the demand for mortgages is somewhat elastic, with an average own-interest rate elasticity of -1.25 . The results confirm that demand side heterogeneity is important. All interaction effects with household income are statistically significant at conventional levels. Heterogeneity is particularly important for households' sensitivity to interest rates. The estimated parameter is three times as large for a household with an income one standard deviation above the average than for a household with an average income. As expected, households' sensitivity to interest rates decreases in their income.

In addition, the estimates imply that there are significant switching costs. For example, a household on average only switches to a mortgage giving it the same utility

1. PRICE DISCRIMINATION, SWITCHING COSTS AND WELFARE

Table 1.6: Structure of banks' policy functions

	Bank A	Bank B	Bank C	Bank D	Bank E	Bank F
Cost of funding	✓	✓	✓	✓	✓	✓
Market share Bank A	✓	✓	✓	✓	✓	✓
- Market share Bank A among richest 50%						
- Market share Bank A among poorest 50%						
Market share Bank B	✓	✓		✓		✓
- Market share Bank B among richest 50%						
- Market share Bank B among poorest 50%						
Market share Bank C	✓	✓				✓
- Market share Bank C among richest 50%						
- Market share Bank C among poorest 50%						
Market share Bank D						
- Market share Bank D among richest 50%						
- Market share Bank D among poorest 50%						
Market share Bank E						
- Market share Bank E among richest 50%						
- Market share Bank E among poorest 50%						
Market share Bank F	✓	✓		✓		✓
- Market share Bank F among richest 50%						
- Market share Bank F among poorest 50%						
R^2	0.911	0.777	0.709	0.823	0.781	0.807

Note: A check mark indicates that the policy function of a bank depends (possibly non-linearly) on the variable. All other market shares conditional on being in a certain income bracket drop out.

if its interest rate is at least 3.4 percentage points lower. This large estimate has two explanations. First, as I argued in Section 1.5, these switching costs measure many different frictions that I cannot differentiate in the data. For example, it is known that in this market there is significant inattention as well: survey evidence indicates that 60% of households that receive an offer to renew their mortgage do not consider switching.⁵⁶ This fact is also picked up by the high switching cost estimate. Note moreover that since both the coefficient on the interest rate and switching costs differ by household income, there is significant heterogeneity in the propensity to switch. For example, a household with an income just one standard deviation below the average switches to an otherwise equivalent mortgage when the interest rate is 1.4 (instead of 3.4) points lower.

⁵⁶Vereniging Eigen Huis. "Huiseigenaren laten massaal geld liggen bij nieuwe rente hypotheek", January 3 2018. <https://www.eigenhuis.nl/actueel/nieuws/2018/01/03/09/00/huiseigenaren-laten-massaal-geld-liggen-bij-nieuwe-rente-hypotheek>

Table 1.7: Demand model estimates

	Constant term		Interaction with household income	
	Coefficient	Std. error	Coefficient	Std. error
Bank				
Bank A	-0.342	(0.008)	-0.888	(0.004)
Bank B	-0.220	(0.009)	-0.721	(0.004)
Bank C	-1.866	(0.012)	-0.610	(0.005)
Bank D	-0.841	(0.009)	-0.868	(0.004)
Bank E	-0.196	(0.010)	-0.869	(0.004)
Bank F	-0.435	(0.010)	-0.898	(0.004)
Payment type				
Annuity	–	–	–	–
Linear	-2.641	(0.002)	0.327	(0.001)
Bullet	0.009	(0.017)	0.325	(0.003)
Savings mortgage	0.311	(0.019)	0.331	(0.004)
Life mortgage	-0.195	(0.014)	-0.256	(0.004)
Investment mortgage	-1.569	(0.014)	-0.102	(0.003)
Duration fixed interest rate				
5 years	–	–	–	–
10 years	2.042	(0.006)	-0.338	(0.003)
15 years	-0.395	(0.012)	-0.440	(0.003)
20 years	0.600	(0.014)	-0.280	(0.003)
Interest rate				
Π_α	-0.423	(0.006)	-0.825	(0.002)
Switching costs				
Π_c	0.807	(0.003)	0.008	(0.001)
Control functions	✓			
Bank dummies	✓			
Log likelihood	-101387237			
Observations	283396			

Note: The table reports the coefficients of households' utility specification (1.1). The baseline product is a 5 year annuity. Standard errors, in parentheses, account for first stage estimation error (Karaca-Mandic and Train 2003). Household income is standardized.

1. PRICE DISCRIMINATION, SWITCHING COSTS AND WELFARE

Table 1.8: Average marginal costs by payment method and fixed interest rate duration

	Dynamic				Static			
	5 yr.	10 yr.	15 yr.	20 yr.	5 yr.	10 yr.	15 yr.	20 yr.
Annuity	0.87	1.49	1.82	1.89	0.87	1.42	1.80	1.83
Linear	0.64	1.26	1.59	1.66	0.66	1.21	1.59	1.62
Bullet	0.49	1.11	1.44	1.51	0.48	1.03	1.41	1.44
Savings	0.52	1.14	1.47	1.54	0.51	1.07	1.44	1.47
Life	0.97	1.59	1.91	1.99	0.94	1.49	1.87	1.90
Investment	0.76	1.38	1.70	1.78	0.78	1.33	1.70	1.74

Note: The table shows average marginal costs per mortgage type. The average is unweighed over the banks in my sample. Marginal costs are calculated for an interest rate on Dutch government bonds of $i_t = 1.37$, the average in the post-ban period. The left-hand side of the table shows marginal cost estimates for the model described in the main text. The right-hand side of the table shows marginal cost estimates under the assumption that banks are not forward-looking.

1.7.1.3 Marginal cost estimates

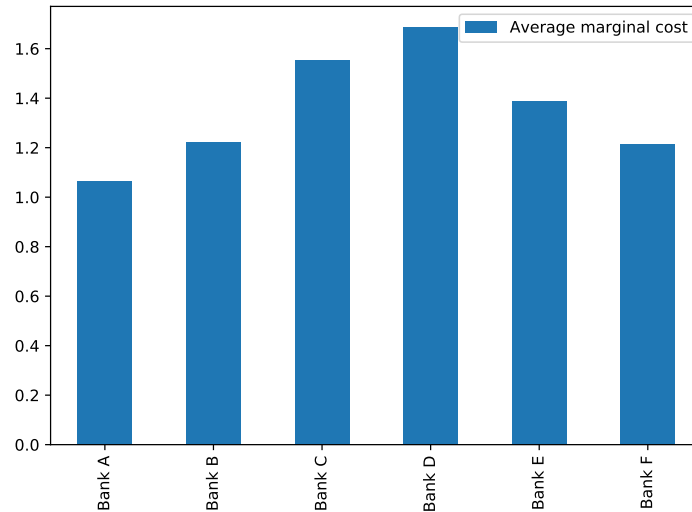
Table 1.8 contains the average marginal costs of the different types of mortgages sold in the Dutch mortgage market. For comparison, I also estimate the supply model under the assumption that banks are not forward-looking and optimize their flow profits. On the whole, the patterns make sense. Longer fixed interest rate periods are associated with higher marginal costs. The marginal costs of bullet, savings and investment mortgages are lower than of other types of loans. This reflects that these type of mortgages cross-subsidize other (often mandatory) products sold by the same bank, such as investment products. Finally, the interest rate on demand deposits, an important source of funding for Dutch banks, was around .5 percentage points in this period. This is consistent with the estimated marginal costs for 5-year loans, which are only somewhat higher.

The estimates of marginal costs are on the whole somewhat larger under the assumption that banks are forward-looking under static maximization. For the most common fixed interest rate duration, ten years, the difference is about 5%. This means that would I not account for banks' dynamic incentives, I would overestimate banks' profits.⁵⁷

Figure 1.4 shows the dispersion of marginal costs across banks. Because the dispersion is significant, there is potential for history-based price discrimination to introduce cross-segment inefficiencies—consumers purchasing from inefficient banks because they are locked-in there. An important mechanism to keep in mind when evaluating

⁵⁷This mirrors the recent finding of MacKay and Remer (2018) that it is important to account for forward-looking behavior of firms in markets with customer inertia: firms have a so-called "investment motive" in that they might want to charge a lower price than under static profit optimization to attract future locked-in customers. Here, in a static setting the interest rates banks set can only be explained by relatively low marginal costs. However, once I account for banks' forward-looking behavior estimated marginal costs are higher. However, the differences I find are not as large as in MacKay and Remer (2018).

Figure 1.4: Marginal costs across banks



Note: The figure shows a histogram of average marginal costs per bank, where the average is taken over all mortgages a bank offers. The marginal costs are calculated for the interest rate on 10-year Dutch bonds equal to $i_t = 1.37$, the average in the post-ban period.

the ban will thus be the (re-)allocation of consumers across banks.

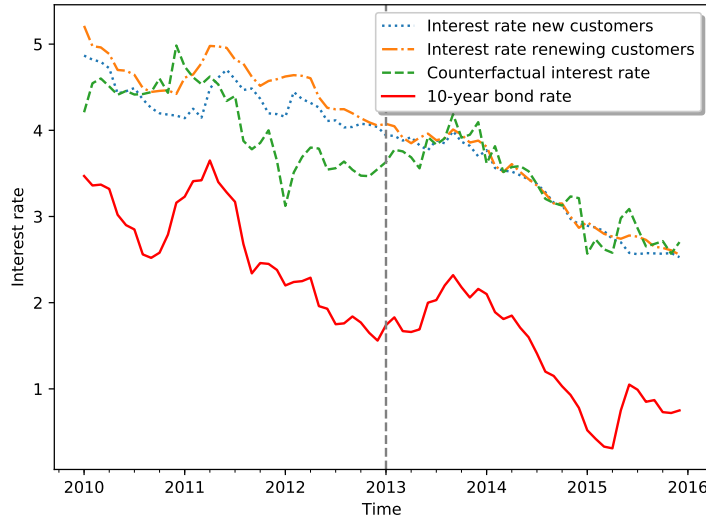
1.7.2 The effect of history-based price discrimination on consumer surplus, profits and welfare

Figure 1.5 shows the predicted average interest rate implied by the estimated policy functions. The post-2013 prediction is in-sample and shows the fit of the model. The pre-2013 prediction is out-of-sample. Therefore, it shows the counterfactual interest rates implied by the model, had the ban on history-based price discrimination been instituted before 2013. The counterfactual uniform interest is typically below the interest rates for renewing and for new customers.⁵⁸ This means that the rent-extraction effect is stronger than the competition effect.

Figure 1.6 shows the counterfactual interest rates when I do not estimate an SMPE and do not use the Lasso, but instead make the ad hoc assumption that banks care about the market shares of all banks, but not how different types of households are distributed amongst them. This is an assumption that in principle seems very reasonable, but Figure 1.6 shows clearly that the resulting counterfactual interest rates are not:

⁵⁸One may expect that the uniform interest rate must lay between those two. However, Corts (1998) shows that in oligopolies third-degree price discrimination may lead to prices for *all* consumer segments to be below or above the uniform price. This can happen when firms have different “strong” markets, i.e. markets where the elasticity of demand they face is comparatively low. When there are switching costs, this the case: a bank’s locked-in customers form its strong market.

Figure 1.5: Counterfactual and observed interest rates



Note: The figure shows observed interest rates and counterfactual interest rates as predicted by the structural model. All are weighed by observed market shares. The interest rate on 10-year Dutch bonds is included for comparison.

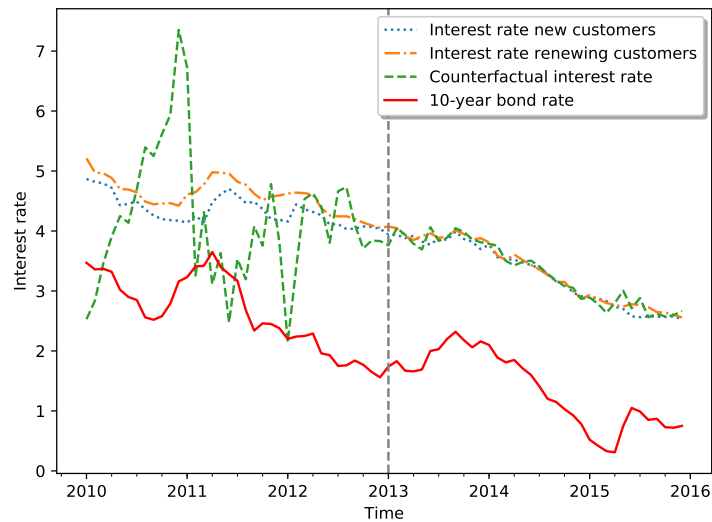
the counterfactual interest rates are erratic and often quite high or low. This shows the danger of ad-hoc assumptions and the importance of data-driven methods: even assumptions that a priori seem reasonable run the risk over over-fitting and can create unrealistic outcomes when extrapolated to out-of-sample states.

Given the counterfactual interest rates, I can calculate the effect of the ban on consumer surplus and bank profits. Table 1.9 shows the estimated effects on consumer surplus. For an average mortgage of €150,000, expected consumer surplus increases by €415 per year. Households whose mortgages are up for renewal gain the most, €588 per year for an average mortgage. This is particularly because the consumer surplus of renewing at their current bank is more attractive, since interest rates are lower under the ban: the expected consumer surplus of renewing increases by €848. Households are also about 3.3 percentage points less likely to switch. Since switching is costly, this means that the consumer surplus of switching also increases, by €434 per year.

While consumers gains, banks lose. Table 1.10 shows that, for an average mortgage of €150,000, the ban on history-based price discrimination causes a loss in profits of €290 per mortgage per year. This loss can be split up into two parts: the first, which I call the reallocation effect, calculates profits under counterfactual market shares but observed interest rates. The reallocation effect thus measures if consumers purchase from more efficient banks because of the ban.⁵⁹ My results indicate that this is the case.

⁵⁹Theory suggests that price discrimination might cause cross-segment inefficiencies (Stole 2007), that is, might cause consumers to purchase from inefficient firms.

Figure 1.6: Counterfactual and observed interest rates when the Lasso is not used



Note: The figure shows observed interest rates and counterfactual interest rates as predicted by the structural model, when the Lasso is not used to estimate policy functions. Instead the ad-hoc assumption is made that policy functions depend on the market shares of all banks, but not on the distribution of household incomes per bank. All are weighed by observed market shares. The interest rate on 10-year Dutch bonds is included for comparison.

Table 1.9: The effect of the ban on history-based price discrimination on consumer surplus

	Potential first-time buyers	Renewing households	Total
Probability of switching (%)		-3.259	
Consumer surplus of switching (€)		430	
Consumer surplus of renewing (€)		848	
Total consumer surplus (€)	185	588	415

Note: The table shows the yearly effect of the ban on history-based price discrimination on consumer surplus for a mortgage with a balance of €150,000. The effects are calculated by evaluating the estimated structural model at observed and counterfactual interest rates, then taking the difference. The effects are an average over the post-ban period 2010-2013, as well as an average over all households who were active in the market during that period, weighed by loan sum.

1. PRICE DISCRIMINATION, SWITCHING COSTS AND WELFARE

Table 1.10: The effect of the ban on history-based price discrimination on bank profits

	Difference (€)
Profits	-290
– Reallocation effect	229
– Price effect	-519

Note: The table shows the yearly effect of the ban on history-based price discrimination on bank profits for a mortgage with a balance of €150,000. The effects are calculated by evaluating the estimated structural model at observed and counterfactual interest rates, then taking the difference. The effects are an average over the post-ban period 2010-2013, as well as an average over all households who were active in the market during that period, weighed by loan sum. The reallocation effect calculates the effect on profits of counterfactual market shares, holding interest rates constant at observed levels. The price effects measures the effect on profits of counterfactual interest rates, holding market shares constant at counterfactual levels.

For an average mortgage of €150,000, average bank profits increase by €229 per year because of consumer reallocation. Of course, this average belies some heterogeneity: more efficient banks gain and less efficient banks lose.

The second effect is the price effect. The price effect measures the change in bank profits going from observed to counterfactual interest rates. Since interest rates on the whole are lower under the ban, the price effect is negative: holding the market shares fixed at the counterfactual market shares, lower interest rates cause banks' profits to decrease with €519 per year for a mortgage of €150,000. As it turns out, the price effect dominates the reallocation effect and the ban on history-based price discrimination causes average bank profits to decrease.

Adding up, the ban on history-based price discrimination causes an increase of total welfare of €125 per year for a mortgage of €150,000. For the whole NHG segment of the market, this implies a welfare increase of about €11.5 million per year.

To summarize, why does the ban on history-based price discrimination increase welfare? As it turns out, it reduces the following three inefficiencies. First, as mentioned above, the average interest rate is lower without than with history-based price discrimination. Because interest rates are closer to marginal costs, efficiency increases. Second, there is less switching without history-based price discrimination. Since switching costs are a deadweight loss, this also increases welfare. Finally, I find that reallocation towards more efficient banks is crucial to understanding the welfare increase: without the reallocation effect, the effect of the ban on welfare would be $€125 - €229 = -€104$. Therefore, the focus of the theoretical literature on history-based price discrimination on symmetric firms is misguided: cost asymmetries and reallocation can be of first-order importance to understanding the effects of history-based price discrimination.

1.8 Discussion

In this section I discuss the robustness of my results. First, I discuss other changes in the Dutch mortgage market around the time of the ban on history-based price discrimination and how those changes might bias my results. Then, I try quantify this potential bias by computing pre-ban interest rates from the post-ban model. Finally, I discuss the empirical content of SMPE and how this can be used to test whether estimated attention vectors are consistent with sparse maximization.

1.8.1 Other changes in the Dutch mortgage market

The estimates of the effects of the ban on history-based price discrimination are causal to the extent that there are no exogenous changes to the Dutch mortgage market that cause interest rates to differ before and after the ban. My estimates control for changes in the cost of funding and previous market shares. However, various other changes happened in the Dutch mortgage market around the ban on history-based price discrimination. I now discuss some of these changes and how they might impact my results.

The general consensus is that after 2013 the Dutch mortgage market became more competitive. The market share of smaller banks grew and margins decreased (Fransman 2017). My results are consistent with this change—I find that the ban on history-based price discrimination made the mortgage market more competitive. Indeed, this was the main aim of the Dutch competition authorities in proposing the ban. However, because I do not have a control market, I cannot definitively rule out that this increase in competitiveness was caused by other factors than the ban. This would imply that my estimates of the effects of the ban on consumer surplus are biased upwards and my estimate of effect of the ban on bank profits is biased downwards.

A second change is the ban on non-amortizing mortgages for new home purchases in 2013. This leads to a change in households' choice sets, which as I describe in Section 1.5, I control for. This change in choice sets could as a secondary effect also change mortgage pricing. If banks' pricing changes as a result of this ban, one would expect the difference in interest rates between amortizing and non-amortizing mortgages to change. I test this implication in Table 1.11, where I regress the difference in interest rates between annuity and bullet loans on loan fixed effects and a dummy indicating the post-ban period. I find no statistically or economically significant change in the difference between the interest rates of annuities and bullets after 2013. Therefore, I conclude that the ban on non-amortizing mortgages for new purchasers after 2013 has no effect on mortgage pricing.

1. PRICE DISCRIMINATION, SWITCHING COSTS AND WELFARE

Table 1.11: The interest rate difference between annuity and bullet mortgages, pre- and post-ban

	(i)	(ii)
After 2013	0.034 (0.018)	-0.115 (0.063)
Bank \times fixed interest rate duration fixed effects	✓	✓
Month fixed effects		✓
Observations	1728	1728

Note: The table displays the change in the interest rate difference between annuities and bullets after the ban on history-based price discrimination. HAC-robust standard errors are in parentheses.

Another potential confounder might be an exogenous change in the demand for mortgages. Given that the economy and housing market markedly improved after 2013, it is natural to assume that, if anything, the demand for mortgages increased during this period. Since this would imply that interest rates increase, this is difficult to reconcile with the general observation that the market became *more* competitive after 2013. However, to the extent that the demand for mortgages increased after 2013, ignoring this in my analysis would only offset the fact the market became more competitive.

A final change to the Dutch mortgage market is that regulations for mortgage brokers became stricter over time. Perhaps the most important change is the 2013 ban on commission payments from banks to brokers. This could have an important effect on the mortgages brokers recommend, and since about half the mortgages are sold through brokers, on the overall market. As Thiel (2018a) shows, banning commissions could lead to lower interest rates as commissions soften competition. Ignoring this could therefore lead to a further overestimation of the welfare improvement of the ban.

1.8.2 Predicting pre-ban interest rates

As a further check on the model, I compute pre-ban interest rates from the estimated post-ban model. The idea is that, if the other changes in the Dutch mortgage market around the time of the ban are inconsequential, I should be able to predict pre-ban interest rates from the post-ban model by allowing banks to engage in history-based price discrimination. I do this by following Hortacsu and Puller (2008): given estimated policy functions of its competitors, I calculate a bank's response and see how close it is to its actual interest rates.

To do so, I first re-estimate banks' policy functions based on the pre-ban data. The reason I do this is that when history-based price discrimination is possible banks may

find it optimal to pay attention to different state variables than when it is not. Using time series cross-validation as in Section 1.6.1.2 gives estimated sparse states with significantly more variables than in the post-ban period (i.e. more than in Table 1.6). Because calculating a bank's best response falls prey to the curse of dimensionality when it pays attention to many variables, I choose a larger penalty parameter than indicated by time series cross-validation in the Lasso estimation to make calculation of best responses feasible.⁶⁰

Then, I solve bank b 's dynamic profit maximization problem. Bank b sets its interest rates in sparse state $\hat{\sigma}_b$, where the interest rates of bank b 's competitors are calculated from their estimated policy functions. That is, for every sparse state $\hat{\sigma}_b$, bank b maximizes

$$\pi_b(\hat{\sigma}_b, r, \hat{r}_{-b}(\hat{\sigma}_b)) + \beta \mathbb{E}[V(\Gamma(\hat{\sigma}_b, r, \hat{r}_{-b}(\hat{\sigma}_b)))]. \quad (1.9)$$

Here, \hat{r}_{-b} are the estimated policy functions of b 's competitors, evaluated in the sparse state $\hat{\sigma}_b$. I calculate a bank's best response using value function iteration. Full details of the algorithm can be found in Appendix 1.B.3. Here, I mention only two important simplifying assumptions I make during this calculation.

First, I assume that banks only sell ten-year mortgages. Table 1.2 shows that almost 70% of mortgages are of this duration. This greatly simplifies the calculation of (1.9), since otherwise the future states at five, fifteen and twenty years also have to be taken into account. Given this assumption, I use a discount factor of 2% per year (around the average of the cost of funding in the post-ban period), so that $\beta = .98^{10}$.

A second simplifying assumption I make is that there is no uncertainty around the cost of funding. Because the estimated policy functions (Figure 1.5) vary almost one-to-one with the cost of funding, I assume that the cost of funding is constant. This reduces the computational time greatly since it removes the need to calculate expectations.

Table 1.12 displays the percentage deviation of the computed best response interest rates from the observed interest rates. The first pattern that emerges is that the model does an adequate job when predicting the interest rates charged to renewing customers: for three out of six banks, the model predicts interest rates within 5% of the observed interest rates, for five out six within 12%. Moreover, for half of the banks the model predicts higher interest rates than observed, for half lower. This suggests that the other changes in the Dutch mortgage market around the time of the ban on history-based price discrimination did not systematically increase or decrease interest rates.

The computed interest rates for new customers show larger deviations from observed interest rates than the computed interest rates for renewing customers. For three

⁶⁰To be precise, I choose the penalty parameter such that every bank pays attention to two or three state variables.

1. PRICE DISCRIMINATION, SWITCHING COSTS AND WELFARE

Table 1.12: Predictions of pre-ban interest rates from the post-ban model

	Bank A	Bank B	Bank C	Bank D	Bank E	Bank F
% deviation interest rate renewing customers	4.985	28.8	-11.395	8.995	-2.305	-0.915
% deviation interest rate new customers	-29.69	3.345	-11.135	-5.275	-41.065	-26.435

Note: The table shows percentage deviations of computed best response interest rates from observed interest rates, averaged over all mortgages a bank sells.

out six banks, the model does an adequate job of predicting observed interest rates. For the remaining three banks, the computed interest rates are significantly lower than observed. There are two possible interpretations of this fact. The first is that the divergence between computed and observed interest rates is caused by the other changes in the Dutch mortgage market around the time of the ban on history-based price discrimination. This is however difficult to reconcile with the size of the divergence and the fact that this divergence only occurs for the interest rates for new customers. If, for example, the very low interest rates computed for bank A's new competitors were due to the increased post-ban competitiveness of the market, it is difficult to explain why A's computed interest rates for renewing customers are higher than observed. A second interpretation is that economic models of history-based price discrimination do not capture the mechanisms of the Dutch mortgage market well. In this interpretation, the calculated interest rates overestimate the importance of the so-called investment motive. The investment motive is the incentive of banks to set a low interest to capture a large market share. My computations predict a much larger incentive motive than observed in the data. This shows the importance of having exogenous variation in pricing: a counterfactual simulation, such as in Cosguner, Chan, and Seetharaman (2016), would imply much lower interest rates for new consumers than in fact observed.

1.8.3 The empirical content of SMPE

In this section, I discuss the empirical content of SMPE. SMPE puts restrictions on the data. These restrictions can be used as a specification test, to choose the Lasso penalty parameter or to estimate the cost of attention.

SMPE has empirical content because under sparse maximization banks pay attention to a state variable if the cost of inattention towards that variable is sufficiently large. Because the cost of inattention can be calculated from an estimated model, this is a testable implication. To be precise, recall that a bank pays attention to state variable σ_i if and only if

$$\Lambda_i \geq \kappa,$$

i.e. when the cost of inattention is greater than the cost of attention. Since I take the default state equal to the long-run average, i.e. $\sigma_i^d = \mathbb{E}[\sigma_i]$, one of the following inequalities must hold for every σ_i :

$$\begin{aligned}\text{Var}[\sigma_i] r_{\sigma_i}^T v_{rr}^b r_{\sigma_i} &\geq \kappa \text{ if } m_i = 1, \\ \text{Var}[\sigma_i] r_{\sigma_i}^T v_{rr}^b r_{\sigma_i} &\leq \kappa \text{ if } m_i = 0,\end{aligned}$$

where

$$v^b(\hat{\sigma}, r) = \pi_b(\hat{\sigma}_b, r, \rho_{-b}(\hat{\sigma}_b)) + \beta \mathbb{E}[V_b(\Gamma(\hat{\sigma}_b, r, \rho_{-b}(\hat{\sigma}_b)))].$$

The testable implication of SMPE is the existence of a scalar κ that solves the system of inequalities for all state variables. Intuitively, an estimated model is only consistent with sparse maximization if there exists a cost of attention that generates the estimated attention vectors.

To implement the test, note that if a bank pays attention to σ_i but not σ_j , it is possible to add their respective inequalities to get

$$\Lambda_i - \Lambda_j \geq 0$$

or

$$\frac{\text{Var}[\sigma_i]}{\text{Var}[\sigma_j]} \geq \frac{v_{rr}^b r_{\sigma_j} r_{\sigma_j}}{v_{rr}^b r_{\sigma_i} r_{\sigma_i}}. \quad (1.10)$$

Because κ has dropped out, this inequality can be calculated from an estimated model. The intuition behind this inequality is that the more a state variable varies, the higher the cost of inattention becomes. Therefore, if a bank pays attention to σ_i but not to σ_j , σ_i must have a relatively large variance.

Equations (1.10) give a system of moment inequalities for every bank. These can be tested using recent results from the econometrics literature, for example the two-step bootstrap of Romano, Shaikh, and Wolf (2014). One application is as a specification test of the model: are the attention vectors the Lasso chooses actually consistent with sparse maximization? This test can also be used to choose the Lasso penalty parameter λ , instead of cross-validation. One would then estimate the model for various values of λ and choose the value that gives a model that is most consistent with sparse maximization, for example the one with the highest p -value.

I perform these two exercises in Appendix 1.A. First, I find that my model is consistent with sparse maximization for three of out six banks. While I am able to reject sparse maximization for the other three banks, the extent of the violation is not large. For those banks, no more than half of the sample moment inequalities (1.10) are violated. Second, I use this test to find the Lasso penalty parameter λ . This results in attention vectors

that are consistent with sparse maximization for every bank. However, choosing λ this way does not lead to a meaningfully different estimates of the effects of the ban on history-based price discrimination.

The empirical content of SMPE has further applications. For example, it can also be used as a check when computing a counterfactual equilibrium using the estimated attention vectors. Calculating a counterfactual SMPE under the assumption that attention vectors do not change is convenient: one benefits from the reduction of the state space, but does not required to re-calculate attention vectors. However, it is possible that in a counterfactual situation the estimated attention vectors are no longer optimal. Whether this is the case can be assessed using the test derived above. Simulating from a computed equilibrium gives an empirical distribution of $\text{Var}[\sigma_i]$. This empirical distribution can be used to implement the two-step bootstrap described above.

Second, a similar test statistic can be inverted to estimate the cost of attention κ . An application of this is the calculation of counterfactual attention vectors. The inequalities

$$\Lambda_i \geq \kappa$$

for the state variables a bank pays attention to (and the reverse inequalities for those it doesn't pay attention to), can likewise be interpreted as moment inequalities. Testing these inequalities for different values of κ gives a confidence interval for the cost of attention. Note however, that under standard asymptotics, in particular holding the number of state variables fixed, κ will only be set identified.

1.9 Conclusion

This paper studied the effects of a ban on history-based price discrimination in the Dutch mortgage market. In this market, households that renew their mortgage paid on average between €228 and €348 more in interest than new customers for the same loan. Such interest rates differences are possible because of high switching costs.

I estimate the effect of the ban by developing a structural model of demand and supply of the Dutch mortgage market. The supply side is a dynamic game with an infinite-dimensional state space. To deal with this, I introduce a new solution concept, Sparse Markov Perfect Equilibrium (SMPE). In an SMPE, banks only pay attention to the most important state variables. I show how these state variables can be identified using the (multitask) Lasso. My results imply that banks pay attention to between two and five variables, these being a common interest rate and the market of share of some, but not all, other banks. Thus, SMPE reduces the dimension of the problem from infinity to a maximum of five.

Estimates of the demand function and marginal costs then allow me to calculate the effect of history-based price discrimination on consumer surplus, bank profits and welfare. Consumer surplus increases with €415 per year for an average mortgage. This is mainly because locked-in renewing households no longer pay higher interest rates than new customers. Because interest rates are lower, bank profits decrease by €290 per year for an average mortgage, despite the fact that the market share of more efficient banks increases. Adding up, this means that welfare increases with €125 per mortgage per year, or about €11.5 million per year for the whole market.

The analysis has shown the value of SMPE in empirical work. For future research, it would be interesting to investigate its theoretical properties further. In particular, it would be useful to devise an algorithm to calculate an SMPE for a given model.

1.A Testing the implications of SMPE

In this appendix, I further describe how the implications of SMPE can be used to test the model. If the estimated model is correct, equation (1.10) holds for every combination σ_i and σ_j such that bank b pays attention to state variable i but not j . Moreover, such a set of inequalities exists for every bank. Since equation (1.10) only contains quantities that can be estimated or that can be computed based on the estimated demand and supply model, it is easy to test whether it holds in the data. To that end, I note that (1.10) can be interpreted as a moment inequality. To test the moment inequalities I use the two-step bootstrap of Romano, Shaikh, and Wolf (2014). Full details, including how to calculate the right-hand side of (1.10), can be found in Appendix 1.B.4.

In Table 1.13, I test whether the attention vectors identified by the multi-task Lasso are consistent with sparse maximization. Although, as I argued in Section 1.6.1, the Lasso identifies banks' attention vectors, this does not imply that the Lasso selects banks' true attention vectors with probability one. There are two reasons for this. The first is that the Lasso only selects the true non-zero coefficients as the sample size goes to infinity (Belloni and Chernozhukov 2013). In a finite sample, the Lasso can select the "wrong" state variables. The second reason is that identification of the attention vectors is dependent on using the correct functional form (or a superset thereof) in the Lasso regression. There are usually multiple, isomorphic, ways to define the pay-off relevant state and the true policy functions may depend on the state in a highly non-linear way. Therefore, it is easy to misspecify the functional form and it is important to check whether the selected model is consistent with sparse maximization. This also shows the practical benefit of using SMPE: in addition to providing a micro-foundation for using

1. PRICE DISCRIMINATION, SWITCHING COSTS AND WELFARE

Table 1.13: Consistency of estimated policy functions with sparse maximization

	p -value	Fraction of violated moment inequalities
Bank A	.270	12.5%
Bank B	.000	25%
Bank C	1.000	0%
Bank D	.000	33%
Bank E	1.000	0%
Bank F	.000	50%

Note: The first column contains p -values of a test that the estimate model is consistent with sparse maximization. The reported p -value is the largest significance level of the two-step bootstrap test of Romano, Shaikh, and Wolf (2014) for which the null cannot be rejected, where the size of the of the first stage bootstrap is set equal to 10% of the overall size of the test. The second column displays the fraction of sample moments for which (1.10) doesn't hold.

the Lasso during the policy function estimation, it provides a method to check whether the statistical selection of attention vectors has a well-defined economic interpretation.

Table 1.13 shows mixed evidence that the estimated model is consistent with sparse maximization. For three of the six banks in my sample, I cannot reject the null that the bank performs sparse maximization at conventional sizes—for two of those banks, it is even impossible to reject this null at *any* size. For the remaining three banks, I can reject the null of sparse maximization at virtually any size: the p -values of the associated tests are very close to zero. However, it should be noted that the test of Romano, Shaikh, and Wolf (2014) rejects the null whenever only one the inequalities (1.10) is sufficiently violated. The test does not provide information on the extent of the violation. Therefore, I provide in the second column of Table 1.13 the fraction of moment inequalities that are violated in my estimated model. As can be seen, the violations are relatively infrequent. On the whole, if a bank pays attention to one state variable but not another, more often than not this is consistent with sparse maximization. The most common violation is that a bank is estimated to not pay attention to its own market share, but that it does pay attention to another bank's market share.

To check the robustness of the model to the fact that the Lasso does not select attention vectors that are fully compatible with sparse maximization, I re-estimate the model choosing the Lasso penalty parameter λ to maximize the p -values of the test of consistency with sparse maximization, instead of through cross-validation. In other words, I choose the penalty parameter to maximize the model's structural interpretation instead of the out-of-sample fit. Since the most common violation of sparse maximization is that banks do not pay attention to their own market share, I impose that banks must pay attention to their own market share. The other variables I include are the

Table 1.14: Comparison of the estimated effect of the ban on history-based price discrimination for different methods of choosing the Lasso penalty parameter

Effect of ban on:	Consumer surplus	Profits	Welfare
Using cross-validation	415	-279	125
Maximizing the model's structural interpretation	385	-248	137

Note: The values are in euro per year for a mortgage of €150,000. The model estimated using cross-validation is the model described in the main text. When maximizing the model's structural interpretation, one instead estimates the model various values of the Lasso penalty λ and chooses the one with the highest p -value on the specification check described in Appendix 1.A.

outcome of the the multi-task Lasso as in Section 1.6.1.2, except that I choose the penalty parameter λ to maximize the p -values of the null that the model is consistent with sparse maximization. Using this procedure, I get p -values of 1 for all banks, that is a model that is perfectly compatible with sparse maximization. Table 1.14 compares the results of this method with the results obtained cross-validation: they are very similar. Thus, it is not necessary to obtain a model that is perfectly consistent with sparse maximization to gain reasonable results.

1.B Estimation details

1.B.1 Matching households over time

To match switching households over time, I employ the following algorithm. First, I discard all households that do not switch between the LLD in year t and year $t + 1$. I can exactly identify these households because every bank uses a unique scheme to identify households over time.⁶¹ For the remaining loans, I calculate the distance between all loans in the old and the new LLD. The distance between loan i in year t and loan s in year $t + 1$ is the norm between its loan and household characteristics, i.e.

$$d(i, j) = ||X_i - X_j||,$$

where X_i are the standardized characteristics of loan i . As characteristics I take the birth year of the primary borrower, the payment type of the loan, the outstanding balance at the moment of switching and the maturity year of the loan. I then assign loan i as being loan j 's previous loan with probability

$$\frac{\exp\{d(i, j)\}}{\sum_k \exp\{d(k, j)\}}'$$

⁶¹The only exception is ABN Amro, which switches its identifying scheme once. I match those mortgages using the same algorithm.

where the sum in the denominator is over all loans in year t that I cannot match based on loan id in year $t + 1$. I further adjust the matching probabilities such that the aggregate probability of switching equals the switching probability I observe in the DHS and such that the correct proportion of loans in year $t + 1$ is not assigned any previous loan, i.e. is a new mortgage.

1.B.2 Supply side first order conditions

The derivatives of bank b 's profits for a product k with respect to its interest rate for old and new customers are

$$\begin{aligned}\frac{\partial \pi_b}{\partial r_{k0}}(r_b, s) &= \sum_{j \in \mathcal{J}_b} \left((r_{k0} - i_t - \gamma_j) \frac{\partial D_j}{\partial r_{k0}}(b, \psi, r_b, \hat{\rho}^b(s)) \right) + D_k(b, \psi, r_b, \hat{\rho}^b(s)) \\ \frac{\partial \pi_b}{\partial r_{k1}}(r_b, s) &= \sum_{j \in \mathcal{J}_b} \sum_{d \neq b} \left((r_{j1} - i_t - \gamma_j) \frac{\partial D_j}{\partial r_{k1}}(d, \psi, r_b, \hat{\rho}^b(s)) \right) + \sum_{d \neq b} D_k(d, \psi, r_b, \hat{\rho}^b(s)).\end{aligned}$$

(When banks can only set a single interest, the derivative of profits with respect to that interest is the sum of these two terms.) The market share of bank $d \neq b$, if it is in bank b 's sparse state representation, evolves as follows

$$\begin{aligned}\frac{\partial \Gamma_d}{\partial r_{k0}}(r_b, s) &\propto \sum_{j \in \mathcal{J}_d} \frac{\partial D_j}{\partial r_{k0}}(b, \psi, r_b, \hat{\rho}^b(s)) \\ \frac{\partial \Gamma_d}{\partial r_{k1}}(r_b, s) &\propto \sum_{j \in \mathcal{J}_d} \sum_{d' \neq b} \frac{\partial D_j}{\partial r_{k1}}(d', \psi, r_b, \hat{\rho}^b(s)).\end{aligned}$$

I ignore a constant that measures the size of tomorrow's market versus today's market since it will be subsumed by the Lagrange multipliers μ .

Substituting these expressions into (1.8) gives that bank b 's first order conditions in month t can be written as

$$\begin{aligned}\sum_{j \in \mathcal{J}_b} \left((r_{k0} - i_t) \frac{\partial D_j}{\partial r_{k0}}(b, \psi, r_b, \hat{\rho}^b(s)) \right) + D_k(b, \psi, r_b, \hat{\rho}^b(s)) &= \\ \sum_{j \in \mathcal{J}_b} \gamma_j \frac{\partial D_j}{\partial r_{k0}}(b, \psi, r_b, \hat{\rho}^b(s)) + \mu_{bt}^T \frac{\partial \hat{\Gamma}}{\partial r_{k0}}, & \\ \sum_{j \in \mathcal{J}_b} \sum_{d \neq b} \left((r_{j1} - i_t) \frac{\partial D_j}{\partial r_{k1}}(d, \psi, r_b, \hat{\rho}^b(s)) \right) + \sum_{d \neq b} D_k(d, \psi, r_b, \hat{\rho}^b(s)) &= \\ \sum_{j \in \mathcal{J}_b} \sum_{d \neq b} \gamma_j \frac{\partial D_j}{\partial r_{k1}}(d, \psi, r_b, \hat{\rho}^b(s)) + \mu_{bt}^T \frac{\partial \hat{\Gamma}}{\partial r_{k1}}. &\end{aligned}$$

The left side of these equations can be calculated given the estimated demand model and bank b 's policy functions. The right side is composed of quantities that can similarly

be calculated and unknown parameters γ_j, μ_{bt} . The right hand side is linear in these parameters, therefore the supply side first order conditions give rise to a linear system of equations.

1.B.3 Calculating a bank's best response

To find the interest rates r that maximize (1.9), I approximate the value function by a complete product of Chebyshev polynomials of degree 4. I use twenty nodes per state variable. I then use the following algorithm to calculate a bank's best response:

1. Initialize $i = 1$ and $V_b(\sigma_0) = \sum_{t=0}^{\infty} \beta^t \pi(\hat{r}(\sigma_t), \sigma_t)$. Here, $\hat{r}(\cdot)$ are the estimated policy functions of all banks and $\sigma_{t+1} = \Gamma(\hat{r}(\sigma_t), \sigma_t)$. That is, I initialize bank b 's value function as the value function it would obtain when all banks would set interest rates according to their estimated policy functions forever.
2. For every node in the basis of the Chebyshev polynomials, $\hat{\sigma}_b$, calculate bank r_i to maximize

$$r_i = \arg \max_r \pi_b(r, \hat{r}_{-b}(\hat{\sigma}_b), \hat{\sigma}_b) + \beta V_i(\Gamma(r, \hat{r}_{-b}(\hat{\sigma}_b), \hat{\sigma}_b)).$$

3. Calculate bank b 's value function

$$V_{i+1}(\hat{\sigma}_b) = \pi_b(r_i, \hat{r}_{-b}(\hat{\sigma}_b), \hat{\sigma}_b) + \beta V_{i+1}(\Gamma(r_i, \hat{r}_{-b}(\hat{\sigma}_b), \hat{\sigma}_b)).$$

As when calculating the costs of inattention (Appendix 1.B.4), this can be done using a contraction mapping.

4. Terminate if

$$\sup \left| \frac{V_{i+1} - V_i}{1 + V_{i+1}} \right| < \epsilon.$$

Otherwise, increase i by one and go to step 2.

1.B.4 Calculating the cost of inattention

Here, I describe how to calculate the cost of inattention required to calculate the critical value of the moment inequalities (1.10). This quantity depends on an estimated model: since the choice of λ influences the demand estimates (through the control function) as well as the marginal cost estimates (through the sparsity structure of banks' states), both the demand and supply model have to be calculated for each value of λ . Given the model, perform the following steps to calculate $r_{\sigma_i}^T v_{rr} r_{\sigma_i}$ for a given bank b

1. **Calculate the objective value function.** To calculate the test, the objective value function is required. The objective value function measures the actual profits a bank makes given its actions. That is,

$$V_b(\sigma) = \pi_b(\rho_b(\sigma), \rho_{-b}(\sigma), \sigma) + \beta \mathbb{E}_{i'} [V_b(\Gamma(\sigma, \rho_b \sigma, \rho_{-b}(\sigma)), i')].$$

Given the policy functions $\rho(\cdot)$, the value functions be calculated using the following mapping T :

$$V_b = T(V_b) = \pi_b(\rho_b(\sigma), \rho_{-b}(\sigma), \sigma) + \beta \mathbb{E}_{i'} [V_b(\Gamma(\sigma, \rho_b \sigma, \rho_{-b}(\sigma)), i')].$$

Gabaix (2017, Lemma 3.6) implies this is a monotone contraction, so calculating this is easy.

To implement this, I approximate $V_b(\sigma)$ by a complete product of Chebyshev polynomials of degree four on a grid consisting of the Cartesian product of twenty Chebyshev nodes per state variable. The flow profits $\pi_b(\cdot)$ can be calculated using the estimated demand, marginal costs and policy functions.

Denote

$$v_b(r_b, r_{-b}, \sigma) = \pi_b(r_b, r_{-b}, \sigma) + \beta V_b(\Gamma(r, r_{-b}, \sigma)).$$

2. **Calculate the default action.** Recall that the default action contains the interest rates a bank sets in the default state. In the default state, the bank assumes every state variable equals its long-run average in every period. The default action of bank b is

$$r_b^d = \arg \max_r v_b(r, \rho_{-b}(\sigma^d), \sigma^d).$$

The actions of b 's competitors can be calculated using the estimated policy functions.

3. **Calculate r_{σ_i} and u_{rr} .** r_{σ_i} is the derivative of b 's optimal interest rate with respect to state variable σ_i . Since the first order condition is

$$(r - c) \frac{\partial D}{\partial r} + D + \beta \frac{\partial V(\Gamma)}{\partial \sigma} \frac{\partial \Gamma}{\partial r} = 0,$$

the implicit function theorem gives that

$$\frac{\partial r}{\partial \sigma} = (r - c) \frac{\partial^2 D}{\partial r \partial \sigma} + \frac{\partial D}{\partial \sigma} + \beta \frac{\partial V(\Gamma)}{\partial \sigma} \frac{\partial^2 \Gamma}{\partial r \partial \sigma} + \beta \frac{\partial \Gamma}{\partial r} \frac{\partial^2 V(\Gamma)}{\partial \sigma^2} \frac{\partial \Gamma}{\partial r}.$$

(I suppress arguments and the subscript b for legibility.) Simply differentiating the definition of $v(\cdot)$ above twice gives

$$v_{rr} = \frac{\partial^2 r}{\partial r^2} + \beta \frac{\partial \Gamma}{\partial r} \frac{\partial^2 V(\Gamma)}{\partial \sigma^2} \frac{\partial \Gamma}{\partial r} + \beta \frac{\partial V(\Gamma)}{\partial \sigma} \frac{\partial^2 \Gamma}{\partial r^2}.$$

Note that both r_{σ_i} and v_{rr} are evaluated at the default action.

Competition and dynamic pricing in markets with consumer switching costs

2.1 Introduction

How do firms respond to switching costs? At least since Klemperer (1987b, 1987c), the dominant view is that firms' key trade-off is between charging a high price to exploit locked-in customers (harvesting) and charging a low price to expand their market share (investment). This *investment-harvesting trade-off* has been a central ingredient to the subsequent theoretical literature and a key assumption of recent structural empirical work on switching costs.¹ However, there is little evidence that this trade-off is empirically relevant. In other words: do firms in fact price according to this trade-off or are other mechanisms more important in practice?

This paper provides reduced-form evidence that the investment-harvesting trade-off is empirically relevant. It does this by making the observation that dynamic pricing under switching costs provides testable implications on a post-entry equilibrium. To be precise, I show using the model of Somaini and Einav (2013) that entry amplifies the investment-harvesting trade-off. Hence, the relevance of this trade-off can be tested in a market where entry has taken place. I then test these implications using plausibly exogenous entry into the Dutch mortgage market. Overall, my results are consistent with the theoretical predictions.

The first contribution of this paper is to theoretically clarify the effects of entry on incumbents' prices in markets with switching costs. The theoretical literature has

¹Recent empirical papers on firm behavior under switching costs (Cosguner, Chan, and Seetharaman 2016; Rickert 2016; Fleitas 2017; Janssen 2018; MacKay and Remer 2018; Thiel 2018b) all have the investment-harvesting trade-off baked into their models. I discuss the theoretical literature in more detail below.

2. COMPETITION AND DYNAMIC PRICING

been divided on this issue. Some authors, most notably Farrell and Shapiro (1988), predict that an incumbent firm sells exclusively to its existing customers after entry and therefore charges a very high price.² On the other extreme, Klemperer (1989) says that, in markets with switching costs, entry can even lead to price wars. An important limitation of the existing literature is that it has focused on models where the incumbent is a monopolist pre-entry. Therefore, the literature has implicitly restricted itself to situations in which the incumbent is always dominant. However, an increase in competition might have very different effects on dominant than on fringe firms. Therefore, I study the effect of entry in the model of Somaini and Einav (2013). Even though their model has many of the features of the recent empirical literature on firm behavior under switching costs (an arbitrary number of firms with market power, product differentiation, dynamic price competition), it nevertheless has a sufficiently simple structure to elucidate the effects of competition on prices.

Simulating a large number of equilibria from the model developed Somaini and Einav (2013), I show that entry has an ambiguous effect on incumbents' prices. Rather, entry *amplifies* dynamic pricing incentives. That is, for firms with relatively large existing customer bases and therefore a strong harvesting motive, the harvesting motive becomes even more important post-entry. Hence, they increase their prices. On the other hand, for relatively small firms the investment motive is already strong pre-entry. For these firms, entry causes investment in market share to become even more important and they decrease their prices. Moreover, I show that the larger the pre-existing market share of a firm that increases its price, the larger its price increase. Similarly, a firm that decreases its price, decreases it more the smaller its market share. Hence, I show that investment-harvesting dynamics lead to testable implications.

The second contribution of this paper is to empirically estimate the effect of entry on incumbents' prices. I do this in the context of the Dutch mortgage market. In this market, households tend to fix their interest rates, most commonly for a period of ten years. After this fixed interest rate period has ended, a household's current bank offers it to renew its mortgage, that is to fix its interest rate again. At this point a household can also switch to a different bank. However, switching is costly: the largest mortgage broker in the country reports that on average a switching household incurs about €3,700 in fees.³ In addition, switching takes time and effort. Thiel (2018b) estimates switching

²Rosenthal (1980) is another early reference noting that competition can increase prices when some consumers are locked-in.

³De Hypotheker. <https://www.hypothekeer.nl/jouw-woonsituatie/hypotheek-oversluiten/>. Accessed Mar 11, 2019. The broker estimates notary costs of €750, taxation costs of €500, costs of removing the original mortgage from the official mortgage registry of €150, costs for government insurance of the mortgage of €2,000 and advisory costs of €2,990. This comes to a total of €6,390. These costs

costs in this market and finds that they indeed are substantial.

In 2014, various Dutch pension funds entered the mortgage market because of regulatory changes. As shown by Kim and Mastrogiacono (2019), pension funds that were affected more by a change in pension regulations were more likely to enter the Dutch mortgage market. Therefore, this entry can be seen as plausibly exogenous and I exploit it to empirically assess the effects of entry on incumbents' interest rates. For regulatory reasons, the entrants sell a subset of the types of mortgage offered by the incumbents. They sell only *simple* mortgages, which I give this name because they are not sold together with another product, such as life insurance. Incumbents in addition also offer *complex* mortgages, which are sold as a bundle with another product. I argue and give statistical evidence that there is not much overlap between consumers who would be interested in simple rather than complex mortgages. Therefore, the entry essentially splits the market into two: a treatment market, in which entry has taken place, and a control market, in which it has not. Because incumbents face similar demand and cost shocks in the control market as in the treatment market, I can use a difference-in-differences approach to estimate the effect of entry on incumbents' interest rates.

My empirical results are consistent with the theoretical model. Incumbents with small past market shares tend to decrease their post-entry interest rates more than incumbents with larger past market shares. Small incumbents on average decrease their interest rates by 5%, while larger incumbents on average decrease their interest rates by only 2.5%. Zooming in, while there is a clear association between market shares and the response to entry in my data, my results do not completely fit with the theoretical story. In particular, while the association between past market shares and the response to entry is quite strong for a subsample of the smallest banks in the market, the largest three incumbents do not decrease their price less than the largest of those smallest banks (as predicted). Also, I do not observe any incumbent increasing its interest rate post-entry. While this does not contradict my theoretical predictions, the results would nevertheless have given stronger support to the theory had some incumbent increased its price. Hence, while my results are consistent with the investment-harvesting trade-off, they are not strong enough to definitely rule out other explanations.

My results are consistent with firms pricing according to an investment-harvesting trade-off. While there is a large theoretical literature built on this assumption, there is not much work that shows firms actually price according to this theory. Existing papers (Stango 2002; Carbo-Valverde, Hannan, and Rodriguez-Fernandez 2011) essentially rely

are however tax-deductible. Since the average marginal tax rate is 42%, this means monetary switching costs are around €3,700.

2. COMPETITION AND DYNAMIC PRICING

on regressing prices on firms' past market shares. However, these estimates could be biased if there are unobserved shocks to product quality, since such shocks create a correlation between past market shares and prices even absent switching costs.⁴ The approaches of Stango (2002) and Carbo-Valverde, Hannan, and Rodriguez-Fernandez (2011) have the advantage that they directly test the investment-harvesting trade-off, while I only test a derived implication of this trade-off. On the other hand, I view the availability of a control market as a significant advantage of my setting. As a result, I can control for unobserved shocks to product quality.

The theoretical literature on switching costs (Klemperer 1987b; Farrell and Shapiro 1988; Beggs and Klemperer 1992; Padilla 1995; To 1996; Doganoglu 2010; Arie and Grieco 2014; Rhodes 2014; Fabra and García 2015; Cabral 2016; Pearcy 2016; Ruiz-Aliseda 2016) has mostly focused on the competitive effects of the *level* of switching costs, but has not looked at the effects of competition in markets where switching costs are relevant. The aforementioned Farrell and Shapiro (1988) and Klemperer (1989), as well as Klemperer (1987a, 1988), are an exception. Farrell and Shapiro (1988) develop a model in which an incumbent monopolist always sets its price before the possible entrant.⁵ The fact that pricing is not simultaneous might have unintended consequences for the resulting equilibrium (Padilla 1995). Klemperer (1987a, 1988, 1989) all consider finite-period models. While this kind of model can help in elucidating key mechanisms, it is now well known that they might distort the relative importance of the harvesting and investing motives due to game-beginning and game-ending effects (Cabral 2016). I build on the model of Somaini and Einav (2013), which features simultaneous pricing and an infinite horizon. I show that in this framework, the contradicting results of the previous literature can be reconciled: entry can cause incumbents to both increase and decrease their prices.

This paper also contributes to a rich empirical literature documenting the effect of competition on prices. This issue has been studied in the various contexts. Bresnahan and Reiss (1991) look at the effect of competition on prices for a wide array of industries. In retail markets Basker (2005), Hausman and Leibtag (2007), Jia (2008), Basker and Noel (2009), and Ailawadi et al. (2010) focus specifically on the effects of Walmart's entry into a market on prices. The focus of Lira, Rivero, and Vergara (2007) is on the Chilean

⁴Say firm i experiences a positive and permanent demand shock at time t . Then, because the shock causes expansion of demand at time t , firm i 's number of locked-in customers will be relatively high at time $t + 1$. At the same time, firm i will charge a higher price in periods $\tau \geq t$ than before. Hence, unobserved shocks to product quality create a positive correlation between past market shares and prices even absent switching costs.

⁵Biglaiser, Crémer, and Dobos (2013, 2016) also study switching costs in a market where entry can take place, but their focus is more on the effects of switching cost heterogeneity than on the effects of entry per se.

supermarket sector. Abe and Kawaguchi (2010) consider the Japanese supermarket sector. In the US, Courtemanche and Carden (2014) look at the effect of Costco and Sam's Club, while Hosken, Olson, and Smith (2018) study the US supermarket sector more broadly. The effect of competition on prices has also been studied for search markets (Allen, Clark, and Houde 2014; Lach and Moraga-González 2017) and in developing countries (Atkin, Faber, and Gonzalez-Navarro 2018; Bennett and Yin 2019; Busso and Galiani 2019). With the exception of the papers on search markets, this literature has considered only frictionless markets. I add to this literature by studying the effect of competition on prices in markets with switching costs. My theoretical framework gives reasons to believe that the effect of competition might be different in such markets than in frictionless markets. Since switching costs have been widely documented (e.g Handel 2013; Honka 2014; Shcherbakov 2016), it is important to extend the existing empirical literature to this setting.

This paper proceeds as follows. Section 2.2 calculates the effects of entry in the model of Somaini and Einav (2013). Section 2.3 gives more background on the Dutch mortgage market and the recent entry into this market. Section 2.4 describes my data set. In Section 2.5 I develop my empirical strategy and test the developed implications. Section 2.6 concludes.

2.2 The effects of entry in a simple model of switching costs

To theoretically investigate the effects of entry in markets with switching costs, I employ the model of Somaini and Einav (2013). This model is relatively rich compared to much of the theoretical literature on switching costs while still being highly tractable. Because I focus on the effects of entry, it is particularly attractive that the model of Somaini and Einav (2013) allows for an arbitrary number of firms. Most of the existing theoretical literature deals with model of duopoly.⁶

I now briefly describe the setup of Somaini and Einav (2013). I do not alter or extend their model but will just point out the implications of entry in their setting. Therefore, I highlight only the model's most important features and refer the reader to the original paper for more details. There are N firms in the market, each selling a horizontally differentiated product. Differentiation is modeled spatially: the products are located at the vertices of the $(N - 1)$ -dimensional simplex. For example, for $N = 3$, the products

⁶Two exceptions are Arie and Grieco (2014) and Percy (2016), who also allow for an arbitrary number of firms. However, their models are significantly less tractable than that of Somaini and Einav (2013). The other theoretical models of markets with switching costs cited in the introduction all feature duopolies.

2. COMPETITION AND DYNAMIC PRICING

are located at the corners of an equilateral triangle. This model of product differentiation extends the Hotelling (1929) linear city model to higher dimensions and is isomorphic to the spokes model of Chen and Riordan (2007).

Consumers live for two periods. At the beginning of every period, the mass of old consumers is normalized to one. In addition, in every period g new, unattached consumers are born. Note that g can be smaller or larger than one, so that the size of the market can grow, shrink or stay constant over time.

Consumers are located uniformly on the $\frac{N(N-1)}{2}$ edges of the simplex. In the first period, a consumer is assigned a location on one of the edges of the simplex with uniform probability. By assumption, a consumer only considers the two firms located on his own edge. To purchase a certain product, a consumer pays the product's price and incurs a linear transportation cost proportional to the distance he has to travel along the simplex' edge. The market is assumed to be covered, so all consumers make a purchase.

In the second period of a consumer's life, his position is redrawn in the following manner. He receives a new, random position on the same edge of the simplex as he inhabited in the first period. The consumer therefore considers the same two products as in the first period. He however faces a switching cost s if he purchases from a different firm than before. Therefore, in the second period the consumer purchases the product that minimizes the sum of the product's price, transportation cost and (if applicable) switching costs.

Consumers are forward-looking and have correct beliefs on firms' pricing strategies in equilibrium. Hence, in the first period, the consumer purchases the product that minimizes his total discounted costs. Consumers employ a discount rate $\delta_c \in [0, 1)$. The model thus allows for myopic consumers.

Firms are infinitely-lived and maximize their discounted payoffs. A firm faces two sources of demand. The first is from old consumers. Because a firm's demand is less elastic when it has more locked-in consumers, the demand from old consumers depends on past market shares $(x_i)_{i=1, \dots, N}$. In addition, a firm sells to the newly-born, unattached consumers. Firms discount future profits with a discount rate r_f . It is assumed that firms' effective discount rate $\delta_f \equiv r_f g$ is smaller than one. While Somaini and Einav (2013) allow for marginal cost differences, I normalize marginal costs to zero to simplify exposition.

Even though this model is rich in various dimensions—it contains product differentiation, an arbitrary number of firms and an infinite horizon—it still has a surprisingly simple Markov Perfect Equilibrium (MPE):⁷

⁷In the formulation in Somaini and Einav (2013) the slope α also differs from firm to firm. Since I

2.2. The effects of entry in a simple model of switching costs

Proposition 1 (Somaini and Einav (2013)). *There exists a Markov Perfect Equilibrium in which every firm has the policy function*

$$p_i^* = \alpha + \beta x_i,$$

where $\alpha, \beta > 0$ are constants that depend on the model's parameters and where x_i is the firm's market share amongst old consumers.

Somaini and Einav (2013) moreover show that this is the unique MPE in parallel linear strategies.⁸ Not only is a firm's optimal strategy linear, it also depends on just its own past market share rather than the full vector $(x_i)_{i=1,\dots,N}$. This is particularly useful when studying entry, because it means that the effective dimension of the game does not depend on the number of active firms. However, because β is the solution to a fourth-degree polynomial (Somaini and Einav 2013, p. 975), it is not possible to derive useful closed-form solutions for α and β .

Before looking at the effects of entry in this model, it is instructive to study the structure of this policy function. The ratio $\frac{\beta}{\alpha}$ essentially measures the strength of dynamic pricing incentives. If this ratio is large, firms with many locked-in customers charge significantly higher prices than firms with lower past market shares. Indeed, the price ratio of two firms can be written as

$$\frac{p_i}{p_j} = \frac{1 + \frac{\beta}{\alpha} x_i}{1 + \frac{\beta}{\alpha} x_j}.$$

When $x_i > x_j$, this ratio is increasing in $\frac{\beta}{\alpha}$. As $\frac{\beta}{\alpha}$ goes to zero, the firms charge the same price even though i has the larger market share. As $\frac{\beta}{\alpha}$ increases, the price difference between firms with smaller and larger past market shares increases. That is, for larger firms the harvesting motive becomes relatively more important when $\frac{\beta}{\alpha}$ increases, while for smaller firms the investment motive increases in relative importance.

I now study the effects of competition on firms' equilibrium strategies. I do so in the following way. I take the established market shares of the incumbent firms $(x_i)_{i=1,\dots,N}$ as given and look at the entry of firm $N + 1$. An entrant does not have locked-in customers, i.e. $x_{N+1} = 0$. This means that I look at the effects of an increase in competition in the period right after the entry has taken place, and not at, for example, the effects on the steady state. The reason is that, empirically, the periods immediately after entry are what I in fact observe.

assume that marginal costs are the same across firms, the policy functions reduce to the form shown here (see Somaini and Einav 2013, eq. 30).

⁸A linear strategy profile is parallel if the coefficient on x_i is the same for all firms.

2. COMPETITION AND DYNAMIC PRICING

Table 2.1: Parameter values for which the model is simulated

	Min value	Max value	Number of points
Switching cost (s)	.01	.99	10
Discount rate consumers (δ_c)	.01	.99	10
Discount rate firms (δ_f)	.01	.99	10
Growth rate market (g)	.025	2.475	10
Number of firms (N)	3, 4, 5, 6, 7, 8, 9, 10, 20, 50, 100		

Note: The unique Markov Perfect Equilibrium in parallel linear strategies is computed for the Cartesian product of the parameter ranges (110,000 equilibria in total). Intermediate points are taken equidistant between the endpoints.

The first proposition follows immediately from observation of firms' policy functions. Because an entrant does not have an established market share, its price after entry is simply $p_i = \alpha$. In other words, a recent entrant only has an investment motive. Hence, it always charges a lower price than the incumbents.

Proposition 2. *An entrant charges the lowest price in the market.*

Because it does not change the structure of firms' policy functions, any effect of entry operates through the coefficients α and β . , it is difficult to analytically derive comparative statics with respect to the number of firms N . Therefore, I proceed with a numerical investigation. I calculate firms' optimal policy functions on a grid of 110,000 different parameter constellations (Table 2.1). Because I take eleven different values for N , this means that I investigate the effect of competition for 10,000 different environments.⁹ For each environment, the following proposition holds true.

Proposition 3. *As the number of firms N increases, the intercept of firms' policy functions α strictly decreases and the slope of firms' policy functions β strictly increases.*

From this proposition it follows immediately that after entry $\frac{\beta}{\alpha}$ increases. In other words, competition increases dynamic pricing incentives. Indeed, from observing the policy function it is obvious that if α decreases and β increases there exists a cutoff market share x' such that firms increase their price if their market share is below x' and increase their price if their market share is above x' .

Proposition 4. *There exists a cut-off market share, such that incumbent firms with a customer base above this share increase their price after entry and incumbent firms with a customer base below this share decrease their price after entry.*

⁹I thank Somaini and Einav (2013) for making their code for calculating MPE's publicly available.

The intuition for this result is as follows. An entrant has no market share and will therefore price aggressively. An incumbent firm must then lower its price to increase its future market share, which decreases its margin on old consumers. Thus, entry increases the opportunity cost of investing in market share. For firms with large established customer bases, this opportunity cost may be so large that it may then become worthwhile to give up on attracting new customers and instead “feed off” existing consumers. For firms with low market shares this strategy is not an option: they have to decrease their price to compete with the aggressive entrant.

Previous results on the effects of entry in markets with switching costs have only considered the case of a single incumbent and found that the incumbent either always increases (Farrell and Shapiro 1988; Klemperer 1988) or always decreases (Klemperer 1989) its price. Here I show that in a richer model, with more than one incumbent, these results can be somewhat reconciled: after entry, typically dominant firms will increase their price while fringe firms will decrease their price.

Because the effects on β are multiplied by the firms’ market shares, it follows that firms with market shares further away from the cut-off market share change their price more after entry.

Proposition 5. *Among incumbent firms that increase their prices after entry, a firm with a larger customer base increases its price more. Among incumbent firms that decrease their prices after entry, a firm with a smaller customer base decreases its price more.*

Prices can increase after entry in markets without switching costs as well (e.g. Janssen and Moraga-González 2004; Chen and Riordan 2007, 2008). What differentiates the predictions under switching costs from these models is the fact that large firms increase their price more than smaller firms. Thus, the main implication of the model is a correlation between previous market shares and post-entry price responses.

2.3 The Dutch mortgage market

I test the results derived in Section 2.2 in the Dutch mortgage market. The Dutch mortgage market is an ideal testing ground for the theory for various reasons. First, the Dutch mortgage market features significant switching costs. Second, in recent years there has been quite some entry. Third, this entry is caused by regulatory changes in other markets and can therefore be seen as plausibly exogenous. I now discuss these points in more detail.

In the Dutch mortgage market, most households fix their interest rates, typically for a period of ten years. However, the total duration of the mortgage is normally longer

2. COMPETITION AND DYNAMIC PRICING

Figure 2.1: Concentration in the Dutch mortgage market



Note: The figure plots the 3-month moving average of the Herfindahl-Hirschman Index (HHI), by start of the fixed interest rate period. Both new and renewed mortgages are included.

Table 2.2: Entry in the Dutch mortgage market in 2014

Firm	Month of entry	Notes
Hypotruster	August 2014	Originally a mortgage broker; started to sell white-label mortgages financed by pension funds.
Tellius	September 2014	Financed by pension funds and insurance companies.
MUNT	November 2014	Financed by pension funds.
Post-entry market share of entrants		5.2%

Note: The table shows all entrants into the Dutch mortgage market in the year 2014. The month of entry is the origination date of the first mortgage in the Loan Level Data by the Dutch Central Bank. The post-entry market share of entrants is the entrants' share of newly originated mortgages in the period Aug 2014-Jul 2016.

than that—most mortgages are paid off in thirty years. When the fixed interest rate period ends, a consumer's current bank is legally required to offer a new fixed interest rate. At this point the consumer can also switch to a different bank without penalty. This does not mean that switching is free however, as various fees must be paid, for example reappraisal and notarial fees. The largest mortgage broker in the country estimates that monetary switching costs are around €3,700.¹⁰ In addition, switching costs time and effort. Thiel (2018b) estimates the total switching costs in a structural model of this market and finds that, even though there is significant heterogeneity in switching costs, on average switching costs are so high that the average household essentially never switches.

A second attractive feature of the institutional context is that it is illegal to engage in price discrimination based on purchasing history. That is, firms in this market are not allowed to charge lower interest rates to new than to renewing customers. This ban was instituted in 2013 to increase the competitiveness of the market. Thiel (2018b) shows that while there was a small amount of non-compliance with this ban, banks indeed charge almost the same interest rate to new and renewing customers. While in many markets with switching costs, one might worry about (unobserved) price discrimination between new and renewing customers, this can therefore be ruled out here.

Historically, the Dutch mortgage market has been reasonably concentrated (Figure 2.1). In particular, the market has been dominated by three large banks: Rabobank, ABN AMRO and ING, who all have market shares of over 20%. Since the start of the financial crisis, there had been no entry and some smaller foreign players even exited the market. This all changed when in 2014 various new players entered. The source of entry was arguably quite surprising. Rather than (foreign) banks entering the market, Dutch pension funds and insurance companies set up new vehicles to sell mortgages. Table 2.2 displays the entrants in this period. Pension funds and insurance companies typically combined through so-called "servicers" that handle marketing and administration of the mortgages. Two mortgages sold by the same entrant can thus be financed by a different institution. However, this process is transparent to the consumer. The entrants were relatively successful, obtaining a combined market share of 5.2% in the post-entry period.

Pension funds and insurance companies seem to have entered the Dutch market because of changes in the regulations governing their balance sheets. For pension funds, the new Financial Assessment Framework ("Financieel Toetsingskader" or FTK in Dutch) and for insurance companies, Solvency II, caused a flight to safe assets. Since repayment rates in the Dutch mortgage market are high, regulators view investments

¹⁰See Footnote 3.

2. COMPETITION AND DYNAMIC PRICING

in the Dutch mortgage market as relatively safe. Kim and Mastrogiacono (2019) show that pension funds that were affected more by the new regulations were more likely to enter the Dutch mortgage market. The Dutch central bank has also attributed the increased activity of non-bank players during this period to changes in regulations (De Nederlandsche Bank 2016). Because the entry was spurred by regulations outside the Dutch mortgage market, it can be seen as plausibly exogenous.

Another defining feature of the Dutch mortgage market concerns the tax deductibility of mortgage interest payments. For this purpose, the tax code distinguishes two types of mortgages. The first consists of amortizing mortgages. These are mortgages where the balance is paid off during the duration of the mortgage. This category includes annuities and linear mortgages. Interest payments on amortizing mortgages are always deductible from income taxes. The second type of mortgage is non-amortizing mortgages. For these mortgages, households only make interest payments—the balance is paid off as a lump sum on the end date of the mortgage. Often, these mortgages are sold together with savings or investments products, the proceeds of which are used to pay off the balance. I call mortgages that are sold together with another product *complex*, while I denote all other (including amortizing) mortgages as *simple*. Interest payments on non-amortizing mortgages are only tax-deductible for households that already had a non-amortizing mortgage before 2013. As a result, first-time buyers only purchase amortizing mortgages. The fact that these are tax-deductible makes them categorically more attractive than non-amortizing mortgages. This fact will be crucial in separating the treatment and control markets, as I explain below.

2.4 Data

The main data source of this article is the Loan Level Data (LLD) of the Dutch Central Bank (DNB). The LLD are a micro-level data set on most mortgages in the Dutch market. It contains 75%–80% of the market, with only some smaller and foreign players missing (Mastrogiacono and Van der Molen 2015). The LLD are reported quarterly, but this study is based on the reports from the fourth wave only so that I in effect have yearly waves. Participating institutions report their full portfolio. Because the entrants are relatively small, they do not routinely report their data to the LLD. However, in the middle of 2016, DNB required *all* institutions in the market to give a one-time snapshot of the outstanding mortgages in their portfolio. Since the entry I focus on happens in 2014, this means that I have information on the most relevant part of the entrants' portfolio: in the first years of their existence.

Note that since institutions report the *stock* of outstanding mortgages in 2016, and

Table 2.3: Market shares (%) of payment types

	Incumbents	Entrants	Total
Simple loans			
Annuity	36.68	70.06	38.17
Linear	3.62	7.46	3.79
Fixed installments	0.09	0.00	0.09
Bullet	33.65	22.48	33.16
Complex loans			
Savings	16.52	0.00	15.79
Life	6.64	0.00	6.34
Investment	1.66	0.00	1.59
Other	1.13	0.00	1.08
Total	100.00	100.00	100.00

Note: The table displays the distribution of payment types across loans, for loans with a fixed interest period starting in the period Jan 2013–Jul 2016. Entrants are defined as institutions selling their first mortgage in the period Aug–Nov 2014. The remaining institutions are labeled incumbents. A loan is complex if it sold as a bundle with another product. Otherwise, it is simple.

virtually all mortgages are active for at least a couple of years, I observe all sales in the years directly prior to 2016. While I do observe prior periods in my data, I start my sample in 2013. The reason is that at the beginning of 2013 many reforms took place in the Dutch mortgage market.¹¹ Therefore, the pre-2013 market might be fundamentally different than the post-2013 market.

Table 2.3 shows the distribution of payment types in the Dutch mortgage market. Many different mortgages are available. Annuity and linear mortgages, which are typical in many other countries, have an overall market share of 42%. Bullet mortgages, where no payments towards the principal are made during the mortgage, have a market share of 33%. Often, bullets are combined with a savings, life or investment product—these types of mortgages have a combined market share of 24%. The distribution of payment types differs in several respects between incumbents and entrants. First, entrants do not sell savings, life or investment mortgages. Second, entrants sell more annuities and fewer bullets. This is because bullets are non-amortizing mortgages. As explained above, non-amortizing mortgages are not attractive for first-buy buyers for tax reasons and are hence only purchased by renewers and switchers. Since entrants do

¹¹For example, the tax deductibility of non-amortizing mortgages changed to the regime described in Section 2.3, the ban on history-based price discrimination described in Section 2.3 took effect in 2013 and there were changes to broker regulations as well.

not have any renewers, it is logical that they sell relatively fewer bullets.

2.5 The effects of entry on interest rates

Section 2.2 shows that dynamic pricing in markets with switching costs provides testable implications on the responses of incumbents to entry. In particular, Propositions 4 and 5 imply that after entry firms with smaller established market shares decrease their price more. Firms with a sufficiently large number of locked-in customers even increase their price. Proposition 6 implies that price dispersion goes up. In this section, I test these implications using the 2014 entry of mortgage sellers funded by pension funds described in Section 2.3.

2.5.1 Main empirical specification

I use a difference-in-differences approach to control for contemporaneous shocks to the mortgage market. Here, I exploit the fact that the entrants do not sell all types of mortgages that are available in the market. As Table 2.3 shows, the entrants do not sell savings, life or investment mortgages. Hence, it is possible to split the market into two segments: a market *affected* by entry comprised of simple mortgages, and a market *unaffected* by entry, comprised of complex mortgages.¹² Because incumbents face similar demand and cost shocks in both markets, the unaffected market serves as a control for the affected market.

In my main specification, I combine a hedonic pricing regression with a difference-in-differences framework:

$$Interest_{ijmt} = \beta_1 Post_t + \beta_2 Post_t \times Entry_m + X'_{ijt} \gamma + \tau_t + \varepsilon_{ijmt}. \quad (2.1)$$

Here, i indexes the consumer, j the mortgage he purchases, m whether the mortgage is sold in the treatment market and t the start month of the fixed interest rate period. $Post_t$ is a dummy indicating whether entry has taken place.¹³ $Entry_m$ is a dummy indicating whether market m is affected by entry or not. τ_t contains month fixed effects. X_{ijt} contains mortgage and consumer variables which might affect the interest rate a consumer pays. I cluster standard errors at the firm level.

I use this hedonic pricing approach instead of mortgage fixed effects for two reasons. First, when there are switching costs, theory predicts that for given characteristics a bank's interest rates are increasing in its past market share. Since past market shares

¹²I ignore fixed installments and other types of mortgages since they have negligible market shares.

¹³I take the first date of entry among all entrants, i.e. August 2014.

Table 2.4: Switching between treatment and control markets

	2013	2014	2015	2016	Total
Proportion switchers	.049 (.006, .165)	.016 (.000, .088)	.017 (.000, .091)	.068 (.030, .130)	.043 (.023, .074)
<i>N</i>	41	61	59	117	278
<i>p</i> -val. no diff. in switching ¹	.545				

¹ *p*-value of Fisher exact test with the null hypothesis that the proportion of switchers is the same in 2013/2014 as in 2015/2016.

Note: The table displays the proportion of households that switches between a mortgage in the control market and a mortgage in the treatment market in the DNB Household Survey. The baseline population is households present in two subsequent waves of the survey, who i) have at least one mortgage in both waves, ii) have at least one mortgage with a fixed interest period starting in the displayed year. A household with one mortgage is a switcher if the mortgage is from a different market in the first wave compared to second wave. A household with more than one mortgage is considered a switcher if it has only mortgages from one market in the first wave and only mortgages from the other market in the second wave. 95% confidence intervals are exact.

change during my sample, fixed effects cannot fully capture mortgage pricing as predicted by theory. Second, mortgage pricing to some extent takes place on the individual level. This is because households differ in risk, which might affect the interest rate they pay, and because some households might negotiate better interest rates in a way that is correlated with observed household characteristics. A hedonic pricing approach, where a mortgage's interest rate potentially depends on the purchasing household's characteristics, can account for that while mortgage fixed effects cannot.

In X_{ijt} I therefore include firm, fixed interest interest duration and payment type fixed effects. These are the variables that are most important in explaining average mortgage pricing at an aggregate level. I also include a firm's past market share amongst consumers that have to renew their mortgage in a given month. To control for differences in risk between households, I do two things. First, I focus on a subsample of mortgages which are insured against default by the Dutch government. From the perspective of the mortgage sellers, these mortgages have very homogeneous risks. Second, I include the loan's loan-to-value ratio as an explanatory variable. In some specifications, I will control for additional loan- and household-level variables, detailed below.

2.5.2 Identification

2.5.2.1 Separation of treatment and control markets

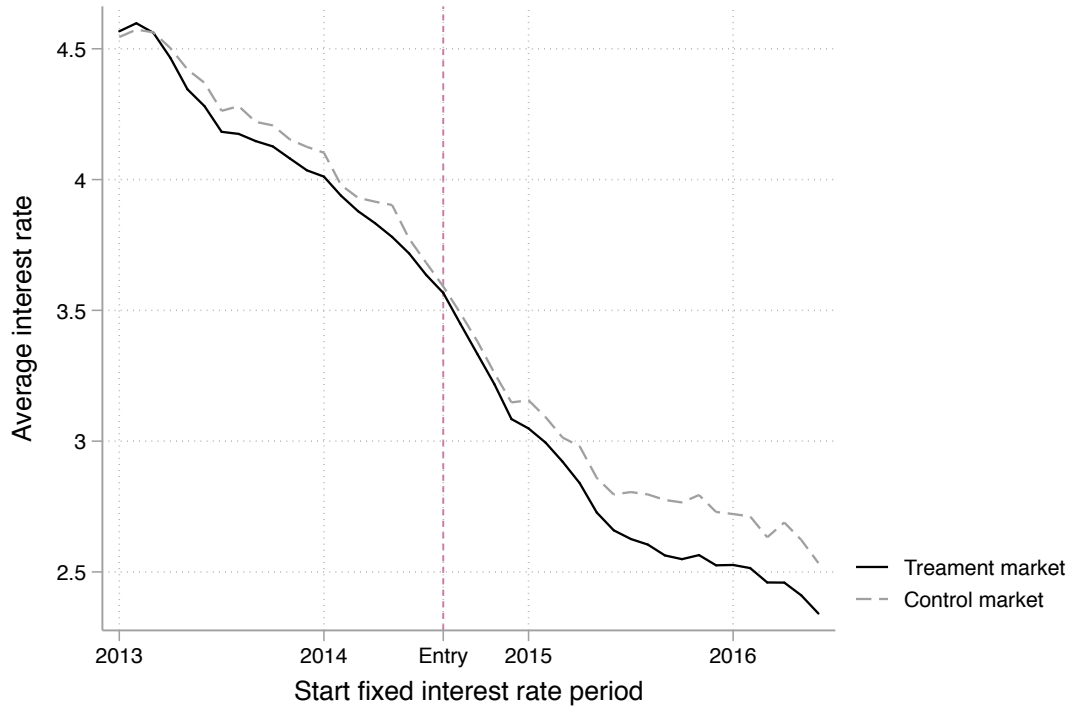
My empirical strategy is valid only if the treatment and control markets can be seen as separate markets. If not, the entry would also affect the control market. That the treatment and control markets can be seen as separate is not immediately obvious, because in principle simple and complex mortgages are substitutes. However, institutional details as well as statistical evidence indicate that to a large extent the treatment and control market can be seen as separate.

The institutional argument is based on the tax treatment of mortgages. As detailed in Section 2.3, from 2013 onwards, first time buyers' interest payments are only tax deductible for amortizing (annuity and linear) mortgages. As a result, first time buyers are only active in the treatment market. The tax advantages of amortizing mortgages are so large that they never substitute towards a complex mortgage. Renewing households can only deduct interest payments on non-amortizing mortgages if they had such a mortgage prior to 2013. Therefore, renewing households similarly never switch from the treatment to the control market.

However, it is possible that renewing households switch from the control to the treatment market. In the LLD the possibility to follow switching households over time is limited. Therefore, I look at the DNB Household Survey to assess how large the fraction of switchers is. The DNB Household Survey is a yearly survey of a random sample of the Dutch population on their finances. Therefore, even though its sample size is much smaller than that of the LLD, it still allows me to obtain a representative estimate of the extent of renewing households' switching. Moreover, it has the advantage that households can be followed from one wave of the survey to the next. Table 2.4 shows that the extent of switching is modest. On average, 4.3% of households that could potentially switch, switch from the control to the treatment market. Moreover, I find no statistical evidence that the rate of switching changed after entry had taken place, as one would expect if the entry had affected the control market.

A second potential challenge to separation is that banks might have cost functions without constant returns to scale. In this case, a bank's sales in the treatment market could affect the costs it faces in the control market. Even if there were no demand-side substitution, this would still imply that banks would change their interest rates in the control market after entry. I do not have information on the structure of banks' cost functions. However, Figure 2.6 shows that most incumbents' sales are relatively constant over time. Hence, even if incumbents' costs feature non-constant returns to scale, it is unlikely that this had a large impact on mortgage pricing.

Figure 2.2: Interest rates over time, by treatment and control market



Note: The figure shows average interest rates of the incumbents by start of the fixed interest rate period for government-backed loans with a 10-year fixed interest rate period.

For these reasons, I argue that the treatment and control market can be seen as largely separate. However, the separation is not perfect. It is unfortunately not obvious what direction a potential bias due to imperfect separation introduces. On the one hand, prices of different firms are strategic complements. Hence, if banks do adjust their interest rates in the control market as a result of the entry into the treatment market, they will do so in the same direction. Therefore, I would underestimate the magnitude of the effect of entry. On the other hand, banks in this market are multiproduct firms. Hence, if consumers substitute between the control and treatment market, banks will internalize the business-stealing effect of a change in interest rates in the treatment market on profits in the control market. This means that they set higher interest rates in the treatment market compared to the case where there is perfect separation and that I underestimate the effect of entry.

2.5.2.2 Parallel trends

My approach identifies a causal effect only if absent entry, incumbents' interest rates in the treatment and control markets would have followed similar trends. This is a plausible assumption, because incumbents face the same cost of funding for mortgages

sold in the treatment and control markets. In the same way, demand shocks, for example due to a marketing campaign, are likely to affect mortgages in both markets similarly. Figure 2.5 plots average interest rates for government-backed loans with a 10-year fixed interest rate duration.¹⁴ The pre-treatment trends are similar: this supports the argument that the treatment and control markets are affected by similar shocks. In Section 2.5.4, I formalize this argument by testing for “leads” and “lags” of the treatment. This exercise also indicates the similarity of the treatment and control markets.

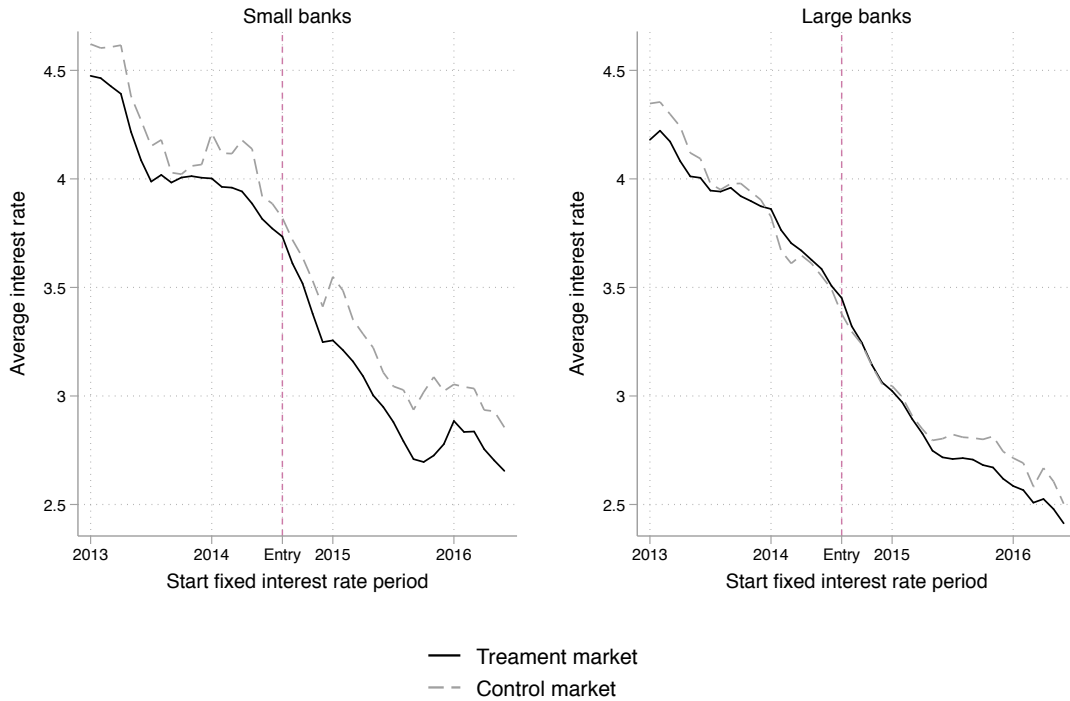
A potential challenge to the parallel trends assumption is that there are systematic differences between borrowers in the treatment and control markets (Table 2.6). Households in the treatment market are on the whole older and richer. They also have smaller loan-to-value ratios and their houses are worth more. Such systematic differences are no problem for difference-in-differences designs *per se*, as long as the parallel trend assumption holds. However, systematic differences between the control and treatment market raise the possibility of there being differential changes in customer composition between the treatment and control market. Such differential changes in composition might render the parallel trends assumption invalid. For example, if younger households on average pay higher interest rates than older households and the number of young households in the treatment market increases relative to the control market, I would overestimate the effect of entry on incumbents’ interest rates. Here, the hedonic pricing approach introduced in Section 2.1 helps. Because I can control for some observable factors that might influence the interest rates that households pay, I can (at least partially) also control for differential changes in composition.

2.5.2.3 Selection and anticipation

In difference-in-differences designs, selection into the treatment is typically a large concern. In this application, however, selection is unlikely to be an issue. This is for two reasons. First, any selection is on the part of entrants, while I only consider outcomes for the incumbents. This means selection is only an issue if entrants select on *incumbents’* post-entry outcomes. While this is certainly not impossible—for example, firms might be more likely to enter in markets where post-entry competition is less fierce—it does limit the amount of selection that is possible. Second, the fact that the entrants sell only simple and no complex mortgages is likely driven by regulation rather than by selection. First, because first-time buyers no longer receive tax benefits for complex mortgages, the control market is slowly dying out. This makes it less attractive to enter this market than the treatment market. Second, a reason why pension funds entered the mortgage

¹⁴I plot a subsample of all loans to make the sample more homogeneous. Government-backed 10-year loans are the most popular category in the market.

Figure 2.3: The effect of entry for small and large incumbents



Note: The figure shows average interest rates by start of the fixed interest rate period for government-backed loans with a 10-year fixed interest rate period by size of the incumbent. An incumbent is considered large if its past market share amongst renewing customers is at least 10% on average in the post-treatment period.

market is because changes in pension regulations made investing in mortgages more attractive. The reason for this is that mortgage debt is seen as relatively safe (Kim and Mastrogiamomo 2019). However, complex mortgages are less safe than simple mortgages. Therefore, it was less attractive for pension funds to enter into the control market. Both these factors are unlikely to be correlated with incumbents' post-entry responses, so that selection is not a large issue here.

A second potential issue is that incumbents might have anticipated pension funds' entry into the Dutch mortgage market. The issue would be that, anticipating fiercer competition for locked-in customers in the future, incumbents would already invest less in market share before the ban. That is, under anticipation effects one would expect the interest rates in the treatment market to increase pre-entry, relative to the control market. Figure 2.2 shows that this does not happen.

2. COMPETITION AND DYNAMIC PRICING

Table 2.5: Main results: the effect of entry on incumbents' interest rates

	(1)	(2)	(3)	(4)	(5)
Market share ¹	0.205 (0.153)	0.207 (0.177)	0.172 (0.219)	0.191 (0.175)	0.0651 (0.169)
Loan-to-value ratio	0.123* (0.0606)	0.122* (0.0596)	0.122* (0.0598)	0.179** (0.0540)	0.119* (0.0587)
Post × Treated	-0.0925** (0.0382)	-0.147** (0.0646)	-0.115** (0.0467)	-0.184* (0.0900)	
Post × Treated × Large bank ²		0.0721 (0.0657)		0.0308 (0.0693)	
Post × Treated × Market share ¹			0.101 (0.216)		
Bank effects	Yes	Yes	Yes	Yes	Yes
Payment type effects	Yes	Yes	Yes	Yes	Yes
Fixed rate duration effects	Yes	Yes	Yes	Yes	Yes
Month effects	Yes	Yes	Yes	Yes	Yes
Additional controls ³	No	No	No	Yes	No
Bank-specific treatment effects	No	No	No	No	Yes
Observations	562679	562679	562679	406495	562679
R ²	0.727	0.728	0.728	0.739	0.731
Joint <i>p</i> -value treatment ⁴		0.065	0.026	0.075	0.000

Standard errors in parentheses

Standard errors are clustered at the firm level

* $p < .10$, ** $p < .05$, *** $p < .001$

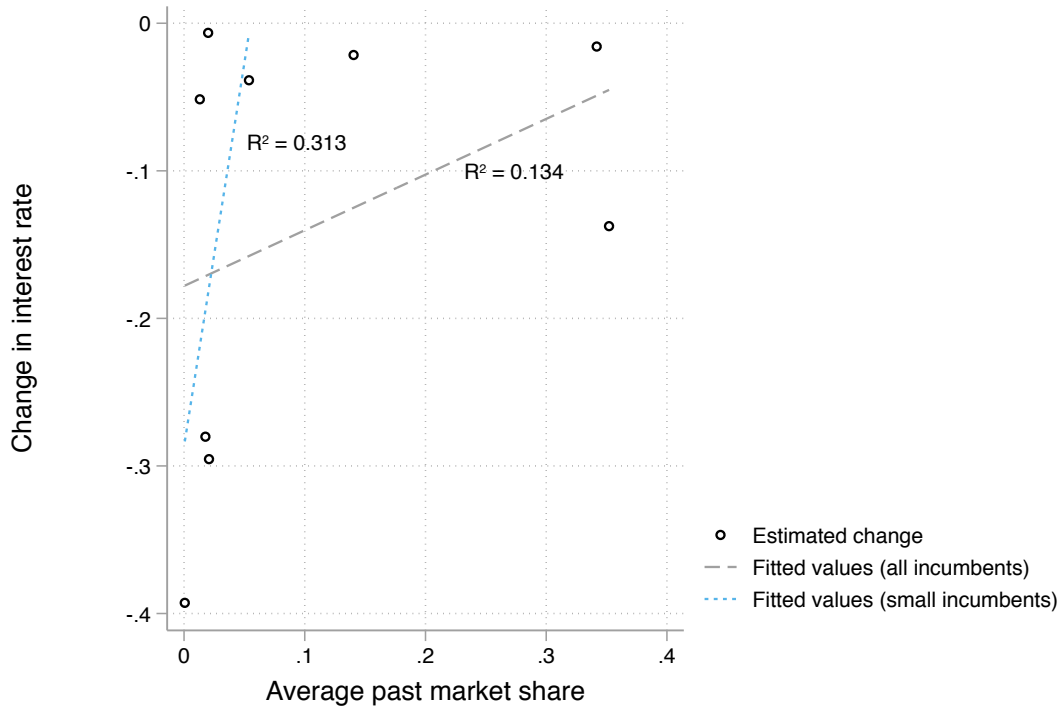
¹ Market share amongst renewing consumers in the relevant month.

² Banks with market shares amongst renewing consumers of at least 10%.

³ Birth year and income of primary borrower; property location fixed effects; original loan balance; mortgage purpose fixed effects.

⁴ *p*-value of the *F*-test of the null hypothesis that there is no treatment effect. For (2) and (4) this null is that "Post × Treated" and "Post × Treated × Large bank" have zero coefficients. For (3) this is that "Post × Treated" and "Post × Treated × Market share" have zero coefficients. For (5) this is that the *Post × Treated × Bank* dummies all have zero coefficients.

Figure 2.4: The response to entry by firm size



Note: The figure displays the scatter plot of past market share and estimated change in interest rate due to entry. The change is estimated by a regression of the form (2.1), where the treatment is allowed to differ by firm. Incumbents with market shares below 1% are removed. An incumbent is small if its average past market share is less than 10%.

2.5.3 Main results

I now estimate the effects of entry on incumbents' interest rates. Figure 2.2 shows that after the entry, average interest rates in the control market decreased relative to the treatment market. Figure 2.3 shows differential interest rate trends for small and large incumbents.¹⁵ It shows that smaller banks, on average, seem to have decreased their rates more post-entry than larger banks. Thus, graphical evidence is consistent with the theory of Section 2.2.

In Table 2.5, I estimate the difference-in-differences specification (2.1). In column (1), I estimate a homogeneous treatment effect. As expected, an increase in competition leads to an overall decrease in interest rates. Entry causes an average decrease in incumbents' interest rates of .09 percentage points, or of 3%.

The remaining columns of Table 2.5 study whether the effect of entry is different for incumbents of different sizes. As predicted by the theory from Section 2.2, larger banks

¹⁵It does not matter whether I classify a bank as big or small based on the treatment, control or full sample.

2. COMPETITION AND DYNAMIC PRICING

decrease their interest rates less than smaller banks. The average interest rate decrease of the three largest incumbents is approximately half that of the other incumbents. Smaller incumbents decrease their interest rate by almost 5%, while this is just 2.5% for larger incumbents. This effect is robust to interacting treatment status with past market shares, as well as inclusion of various other controls.

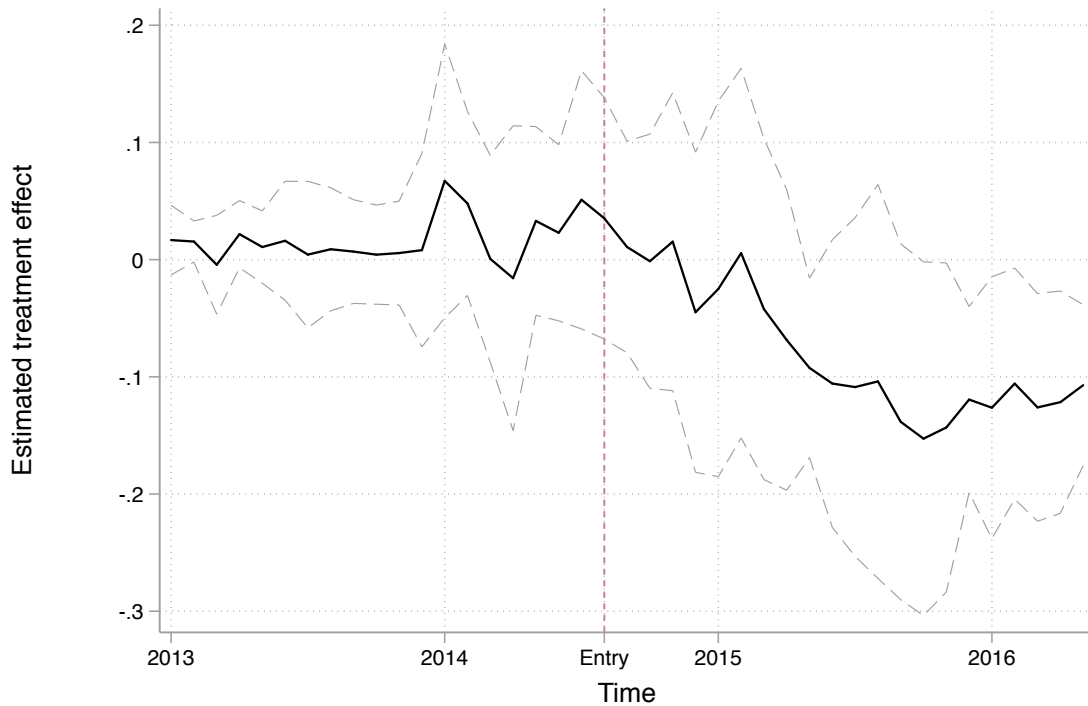
The results are economically significant, but not always statistically significant at standard levels. However, as the last row of Table 2.5 shows, the *joint* effect of treatment is statistically significant. Moreover, it should be noted that I observe the *full* market—not just a random sample of it. As recently pointed out by Manski and Pepper (2018), it is not clear that standard errors should even be included in this case. The reason is that normal standard errors measure the uncertainty due to sampling. When the full population is observed, there is no clear sampling process that leads to uncertainty. Hence, when the full population is observed all that is required is measurement in the population. Nevertheless, I choose to include standard errors to comply with standard practice. In the case where a large (or even complete) fraction of the full population is observed, clustered standard errors are also likely to be needlessly conservative (Abadie et al. 2017a; Abadie et al. 2017b; Athey and Imbens 2018).

In column 5 of Table 2.5 I allow the treatment effect to differ by incumbent firm. Therefore, this is the most flexible specification. Figure 2.4 shows the estimated treatment effect as a function of previous market share. Particularly for smaller incumbents, there is a clear relationship between market share and the reaction to the entry. However, this relationship does not extend as cleanly to the three incumbents. Based on the theory, one would expect the points associated with the largest three incumbents to lay above and to the right of those of the smallest incumbents. While of course the theory is much more stylized than the actual market I investigate, the evidence would be significantly stronger had, for example, the largest incumbents increased their interest rates. Overall, past market shares can explain almost 12% in the response to entry. This provides support for the practical relevance of the investment-harvesting trade-off. Nevertheless, it also leaves a lot of room for other mechanisms to play a role.

2.5.4 Robustness

To test the plausibility of the parallel trends assumption, Figure 2.5 displays treatment trends over time. Consistent with the graphical evidence discussed before, “leads” of the treatment are not significantly different from zero. Figure 2.5 moreover shows that the effect of the treatment increases over time. In other words, it seems to have taken some time for the entry to have full effect on the market. The results should therefore

Figure 2.5: Estimated treatment effects over time



Note: The figure shows the treatment effect with associated 95% confidence interval over time. The per-month treatment effect is the regression coefficient on a month \times treatment group dummy in a hedonic pricing regression. Confidence intervals are based on standard errors clustered at the bank level.

be seen as lower bounds.

2.6 Conclusion

This paper has theoretically and empirically studied the effect of competition in markets with switching costs. On the theoretical front, I study the model of Somaini and Einav (2013). By simulating equilibria of their model for a large number of parameter constellations, I show that, contrary to common intuition for frictionless markets, competition does not need to lead to price decreases. Rather, competition increases dynamic pricing incentives. Firms for which the incentive to invest in market share is already strong, decrease their price post-entry. On the other hand, firms for which the incentive to “harvest” existing locked-in customers is strong, increase their price post-entry.

I test the derived implications by studying entry into the Dutch mortgage market. Because the entrants only sell a subset of all available mortgages in the market, entry splits the market into a treatment and a control market. I employ a difference-in-differences framework, augmented with a hedonic pricing regression, to study the effect

2. COMPETITION AND DYNAMIC PRICING

of entry on incumbents' interest rates. Consistent with theory, I find that smaller banks decrease their interest rates more than larger banks. Nevertheless, past market shares can only explain a relatively small fraction of post-entry interest rate changes. This suggests that perhaps other mechanisms are also important.

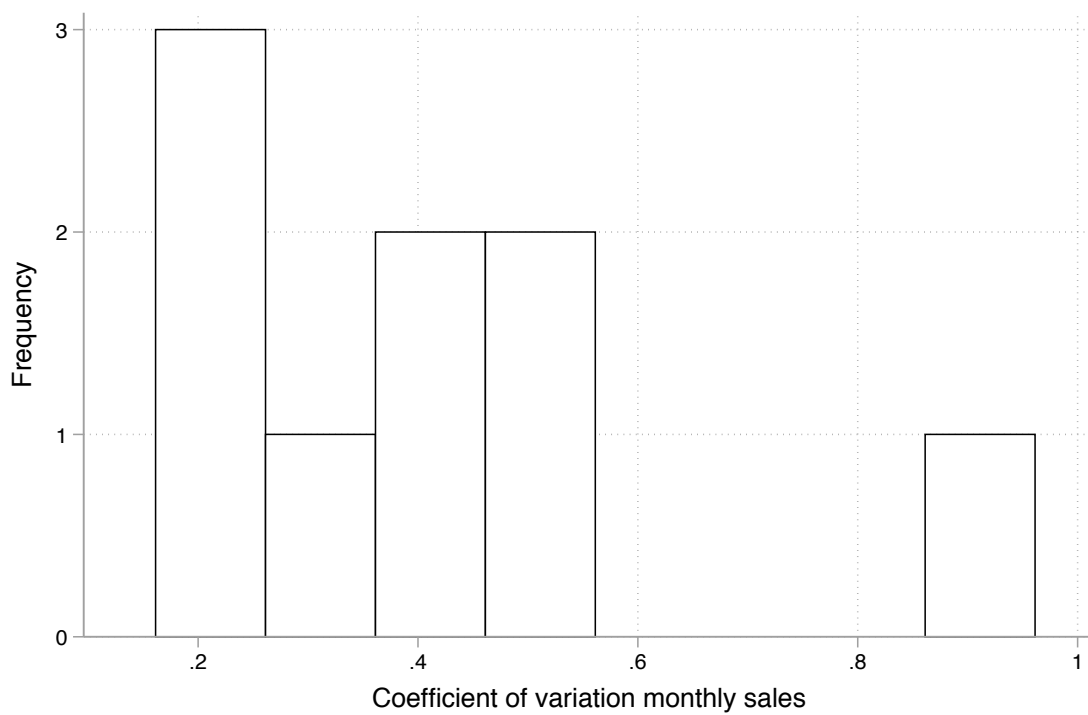
2.A Additional tables and figures

Table 2.6: Borrower and loan characteristics for the treatment and control market

	Control market	Treatment market	Total
Age primary borrower (years)	46.39 (26.26)	50.31 (61.73)	49.43 (55.78)
Household income (€)	60653.3 (64890.8)	62654.9 (97424.2)	62229.5 (91486.6)
Loan-to-value ratio	81.33 (44.26)	74.36 (2825.1)	75.92 (2490.8)
Property valuation (€)	236338.6 (177052.0)	260749.1 (255702.0)	255312.0 (240635.8)

Note: The table displays average borrower and loan characteristics for the treatment and the control market, for the whole sample. Standard deviations are in parentheses.

Figure 2.6: Coefficient of variation of monthly mortgage sales by incumbent



Note: The figure shows the coefficient variation (standard deviation divided by mean) of mortgage sales by incumbents. Sales are measured in total volume of originated or renewed mortgages in euro's. Incumbents with market shares smaller than 1% are not shown.

Fee-based versus commission-based advice

3.1 Introduction

It is common for consumers to consult advisers before purchasing financial products. For example, in 2013, 37.5% of consumers in the United States reported using the services of a financial planner or broker.¹ Historically, financial advisers are remunerated through commissions or kickbacks. In the wake of the financial crisis, regulators around the world became concerned that commissions could cause advisers to give biased advice. Advice is biased when advisers recommend a product with high kickbacks rather than the best product for the consumer. Because of this concern, regulation of adviser remuneration has become increasingly common. For example, the requirement that commission payments have to be disclosed is now almost universal.²

In recent years, some countries, most prominently the United Kingdom, Australia and The Netherlands, have banned commission payments to financial advisers altogether.³ When commissions are banned, advisers charge consumers a fixed fee for advice.⁴ The first experience with commission bans suggests that they are associated with so-called “advice gaps”. An advice gap occurs when some consumers do not have

¹Survey of Consumer Finances, 2013. Accessed from <https://www.federalreserve.gov/econresdata/scf/scfindex.h> on December 5, 2016.

²For example, EU guideline 2014/65/EU requires disclosure of kickbacks in the context of investment products and sets a framework for further regulation by member states. EU guideline 2016/97 regulates kickbacks for insurance advisors. (EUR-Lex, <http://eur-lex.europa.eu/>)

³In the UK and Australia, commissions are banned for investment products. In the Netherlands, commissions are banned for virtually all retail financial products.

⁴The fee is fixed with respect to the recommended product. In practice, the fee can also be hourly or, for investment products, a percentage of the invested sum.

3. FEE-BASED VERSUS COMMISSION-BASED ADVICE

access to financial advice. For example, the Financial Conduct Authority in the UK states that while “the move to fee-based advice on retail investment products has improved transparency and ended conflicts of interest caused by a mainly commission-driven model”, “advice is expensive and is not always cost-effective for consumers” (Financial Conduct Authority 2016, p. 5). Indeed, many consumers report finding fee-based advice too costly.⁵ Thus, there is a trade-off between the quality and the accessibility of advice.

At the same time, advisers claim that fee-based advice puts significant pressure on their margins and harms the sustainability of their business. Taking again the example of the UK, the number of advisers decreased by approximately 25% around the ban on commissions (Chart 1). Such a decline may directly contribute to the advice gap.⁶ In other markets, advisers also report fierce competition after a shift to a fee-based model.⁷ Therefore, a shift to fee-based advice creates a paradox: even though consumers think advice fees are too high, advisers report downward pressure on fees and exit the market.

This paper develops a theoretical framework to explain this paradox and study the trade-off between the quality and accessibility of advice. I study a game in the vein of Inderst and Ottaviani (2012a), in which advisers can recommend one of two horizontally differentiated products. Advisers are remunerated either through kickbacks or directly by consumers. Contrary to the existing literature, I do not take the availability of advice as given. In my model, advisers are capacity constrained: because it takes a fixed amount of time to give a consumer a recommendation, advisers are limited in the number of consumers they can serve.⁸ Moreover, to explain the observed exit of advisers after commission bans, entry and exit of advisers are endogenous.

My main results are as follows. For a fixed number advisers (sufficiently large to serve all consumers), remuneration through kickbacks is typically more efficient

⁵For example, 51% percent of UK households who did not use financial advice but are identified as possibly having a need for it, report being unwilling to “pay for regulated financial advice at any price.”. A further 20% of this group is unwilling to pay more than 100 pounds (Financial Conduct Authority 2017).

⁶That the accessibility of advice is linked to the profitability of advisers is well-understood in the industry. For example, Chris Hannant, director general of the UK Association of Professional Financial Advisers, said: “Only when advice firms’ business models are sustainable and profitable will they be encouraged to enter the market and innovate to help close the advice gap.” From “APFA report shows profits down for financial advisers”. <https://www.wealthadviser.co/2016/03/24/237809/apfa-report-shows-profits-down-financial-advisers>. Accessed October 7, 2018.

⁷Another empirical example of fierce price competition after the shift to a fee-based instead of a commission-based revenue model is the US asset management industry. As noted in a Fidelity Investment report, this shift “[increased] the level of competition in the independent adviser market segment. The report notes that fees have already decreased following a shift to fee-based management and expects further decreases in the future. From the 2016 Fidelity RIA Benchmarking Study. <https://clearingcustody.fidelity.com/app/proxy/content?literatureURL=/9879252.PDF>. Accessed October 7, 2018.

⁸In the UK, the Financial Conduct Authority mentions there not being “enough advisers to go around” an explicit reason for the advice gap (Financial Conduct Authority 2016, p. 34).

than direct remuneration.⁹ The reason, as in Inderst and Ottaviani (2012a), is that kickbacks steer consumers towards the more efficient firm. However, consumer surplus is higher under direct remuneration than under kickbacks. This is because under direct remuneration, advisers always recommend the product giving consumers the highest utility. This finding can explain why regulators, who arguably care more about consumer surplus than efficiency, have banned kickbacks.

When accounting for entry and exit of advisers, this short-run result reverses. Because advice is a homogeneous good, advisers compete their fee towards zero as long as some advisers have spare capacity. Therefore, advisers can only make positive profits if they all operate exactly at capacity. If entry is costly, this implies that in equilibrium the market cannot be covered. Not only does this mean that advice becomes inaccessible to some consumers, but also that competition between advisers essentially disappears: advisers can set fees that extract almost all consumer surplus. On the other hand, under kickbacks sufficiently many advisers always enter the market to serve all consumers. I show that, consumer surplus is always higher under kickbacks than under direct remuneration when accounting for entry and exit of advisers.

These findings can explain the paradox that consumers find fee-based advice costly but that at the same time advisers report significant competitive pressure and exit the market. When switching from commission-based to fee-based advice, the existing capacity is unsustainable. Fees are driven towards zero and advisers exit the market. In the long run, however, sufficiently many advisers exit the market that they regain significant market power and advice becomes inaccessible. Therefore, while banning kickbacks benefits consumers in the short run, it harms them in the long run.

I consider three extensions to the main model. In the first extension, I allow for consumers to be behavioral, in the sense that they systematically undervalue purchasing *any* product. I show that, perhaps contrary to common intuition, behavioral consumers are a reason *not* to ban kickbacks. The reason is that under direct remuneration not all consumers choose to receive advice while under kickbacks they do. Hence, the more they undervalue the value of the product, the larger their ex-post utility.

In the second extension, I allow advisers to recommend the outside option in addition to one of the products. For a fixed number of advisers, this increases the efficiency gains of commission-based over fee-based advice. On the other hand, again for a fixed number of advisers, banning kickbacks is even better for consumers than when advisers must recommend either of the products. However, I argue that my long-run results do not depend on advisers' ability to recommend the outside option.

⁹The only possible exception is when the firms are almost symmetric and advisers put a sufficiently small weight on the utility consumers derive from recommended products.

3. FEE-BASED VERSUS COMMISSION-BASED ADVICE

In the final extension, consumers need not purchase through an adviser but can also determine the best product through costly search. In this case, the fact that consumers have an outside option further lowers the fees advisers can charge. Nevertheless, when search costs are small enough, long-run consumer surplus is higher under direct remuneration than under kickbacks. The reason is that the outside option puts a lower bound on the surplus that advisers leave to consumers: the lower search costs, the higher this lower bound. This implies that policies to reduce the advice gap may have unintended consequences. Only when a sufficient number of consumers can credibly threaten to search for the best match themselves, can banning kickbacks increase consumer surplus. That means that making advice *obligatory*, as in the Dutch mortgage market, is likely counterproductive. However, policies that increase the value of consumers' outside option, such as the encouragement of automated "robo-advice", are likely to benefit consumers.

This paper contributes to the theoretical literature on advice. My main contribution is to build a model of adviser pricing when kickbacks are banned, taking as endogenous the entry decisions of advisers. The previous literature has taken the availability of advice as given. The first experiences with commission bans, as discussed above, however suggest that the endogenous entry and exit of advisers is of first-order importance to understand the effect of commission bans. This paper aims to fill this gap in the literature.

I build on arguably the most important contribution of this literature, Inderst and Ottaviani (2012a). In their model, there is only a single adviser that is always accessible. Therefore, their model cannot capture the shift from commission-based to fee-based revenue models. I embed their model into a model of adviser entry and pricing to study this important policy issue.¹⁰ Another model of the interaction between advisers, consumers and upstream firms can be found in Stoughton, Wu, and Zechner (2011). In their model, the presence of advisers is also exogenous.

Inderst and Ottaviani (2012b) also provide a model of adviser pricing and study the effect of banning kickbacks on consumer surplus. There are however, various differences between their model and mine. Like the other theoretical models on advice, Inderst and Ottaviani (2012b) take the availability of advice as given. Therefore, their model cannot explain the exit of advisers and the reduced availability of advice that have been observed after bans on kickbacks.

¹⁰Although not a main focus of my paper, I further show that their result that the quantity of the low-cost product is always inefficiently low depends on the seemingly innocuous assumption that advisers care about the match value but not about the price of the recommended product. I show that when advisers care also about the price of the recommended product, the quantity of the low-cost firm can also be too high or even efficient.

A second key difference is the type of advice we consider. In Inderst and Ottaviani (2012b), advisers sell a single product from a single upstream firm. In my model, advisers sell products from multiple firms. The model in Inderst and Ottaviani (2012b) can therefore best be understood as one of restricted advice, while my model considers independent advice—the difference being that restricted advisers in effect act as sales agents and only sell products from one firm, while independent advisers sell products from multiple upstream firms. Therefore, the relation between a restricted adviser and a firm is a principal-agent problem, while firms compete for the recommendation of independent advisers. Both types of advice are common in practice. In the UK, for example, 61% of advice is independent while 39% is restricted (Financial Conduct Authority 2016, pp. 18). Hence, my model sheds light on another part of the market for advice than Inderst and Ottaviani (2012b). Bolton, Freixas, and Shapiro (2007) also develop a model of restricted advice.

Whether advice is restricted or independent matters for the effect of a commission ban. For example, Inderst and Ottaviani (2012b) find that banning commissions only increases consumer surplus if a sufficiently large fraction of consumers mistakenly believe that advisers are unbiased. For independent advice, I however find that banning kickbacks increases consumer surplus even if all consumers have rational beliefs on advisers' bias (for a fixed number of advisers). A second difference is that competition between restricted and independent advisers is of a different nature. Independent advisers sell all products in the market and are therefore less differentiated than restricted advisers, who sell differing subsets of all products. This suggests that competition between independent advisers is fiercer than between restricted advisers. Since, in my model, the competition between advisers causes their exit, it is possible that banning kickbacks would lead to a smaller rate of exit for restricted advisers than my model predicts for independent advisers.¹¹

Armstrong and Zhou (2011) also contains a model of commission-based advice. Although not the main focus of their paper, they briefly consider a shift to fee-based advice and argue that this increases consumer surplus. However, in their discussion the fee advisers charge is exogenous—set to equal advisers' commissions. In my model, the market structure is endogenous. I show that, for a fixed number of advisers, consumers are also better off under fee-based advice when advice fees are endogenous. Moreover, I also consider entry and exit of advisers while Armstrong and Zhou (2011) do not.

¹¹While Inderst and Ottaviani (2012b) consider competition between two restricted advisers, each selling a different product produced by a different firm, competition between advisers is modeled in utility space. As a result, it is not immediately clear what their results imply for the effect of banning commissions on the profitability of advisers. Therefore, I cannot make a definitive statement whether banning commissions would lead to more or less exit of advisers in their model than in mine.

There is also some recent empirical work on biased advice. Egan (Forthcoming) studies the impact of commissions on brokers' recommendations in the market for convertible bonds. He empirically confirms the assumption made in theoretical models of advice—including mine—that commissions lead to biased advice. To better understand the role brokers play, he moreover calculates a counterfactual in which brokers are forced to recommend consumers their best option. He finds that consumer surplus increases. My results imply that such a counterfactual does not automatically imply that banning commissions is also a good policy. One has to take into account the entry and exit of advisers if advice is regulated to get a realistic understanding of the regulation's effect. Robles-Garcia (2019) and Myśliwski and Rostom (2018) also find that commissions do bias advice, in this case in the context of the UK mortgage market, but similarly take the presence of advice as given. Another empirical study on advice is Guiso et al. (2018), who study banks' steering in the Italian mortgage market. Their setting is however somewhat different than the one I consider: in their market, banks recommend one of their own products, while I consider independent advisers who recommend a product produced by upstream firms.

Other recent papers on advice relate to other aspects of the market. Bardey et al. (2019) study the incentives advisers have to collect information, while De Cornière and Taylor (2016) consider the incentives of upstream firms to invest in quality in the presence of an intermediary. Shen and Wright (Forthcoming) explain the observation that in many markets with adviser firms do not price discriminate (i.e. there is price coherence) between consumers who purchase through an adviser and those who purchase directly at a firm. Murooka (2015) studies the incentives advisers have to counteract firms' obfuscation strategies.

The remainder of this paper is organized as follows. Section 2 gives an overview of the model I study. Section 3 derives equilibria and studies efficiency and consumer surplus without capacity constraints. In Section 4, I introduce capacity constraints consider entry and exit of advisers and their effect on consumer surplus. Section 5 contains the aforementioned extensions to the main model. Section 6 concludes. Omitted proofs can be found in the Appendix.

3.2 A model of advice

Consumers and firms. A unit mass of consumers looks to purchase a single product from one of two firms. The firms differ in their marginal costs: firm L has marginal costs normalized to zero, while firm H has marginal costs $c \geq 0$. Product valuations are idiosyncratic amongst consumers and firms: consumer k obtains match value ε_{ki} at

firm i , where ε_{ki} is independently and identically drawn from a marginal distribution ε . ε has a log-concave and differentiable density $g(\varepsilon)$ on the support $[0, \bar{\varepsilon}]$, with $\bar{\varepsilon} < \infty$. Denote by $G(\varepsilon)$ the cdf of ε . Consumer k receives utility $\varepsilon_{ki} - p_i$ if he purchases product i and 0 if he purchases no product. Continuing, I will drop the consumer subscript k on ε_i for ease of exposition. Consumers do not know which firm has high marginal costs and which has low marginal costs.

Advisers. To start, consumers require the recommendation of an adviser to purchase a product. In Section 3.5.3, I consider the case where consumers can also purchase from firms directly—in this case, advisers still help consumers to find the best match. The market contains $N \geq 1$ advisers. When a consumer visits an adviser, the adviser observes the match value ε_i and price p_i for both firms. Consumers observe the prices; they observe the match value of a product only after they purchase it. Advisers give a recommendation to purchase one of the two products.¹² An adviser does not reveal any more information, such as the match value of the recommended product.¹³ To start, advisers can serve an unlimited amount of customers and the number of advisers active in the market is fixed. In Section 3.4, I consider capacity-constrained advisers and advisers' entry.

In addition to receiving a recommendation, consumers receive a convenience benefit b_j when they visit adviser j (as in Edelman and Wright 2015b). This convenience benefit can reflect that in many markets advisers do more than recommending the most suitable product—for example they might additionally offer tax planning services. It can also reflect some other aspect of horizontal differentiation between advisers, such as their location. Consumers observe the convenience benefits of all advisers before they choose which adviser they visit. Denote the convenience benefit consumer k obtains from adviser i by b_{ki} . b_{ki} is drawn independently and identically from a distribution $H(b)$ on a support $[0, \bar{b}]$, with $\bar{b} < \infty$. Assume that the density $h(b)$ is log-concave. Continuing, I also drop the consumer subscript k on b_i . Consumers do not need to make a purchase to receive the convenience benefit.

If \bar{b} is sufficiently large, advisers can run a profitable business by only delivering convenience benefits and not giving any advice. I want to focus on the case where advice is the core business of advisers, as in real-world markets. That is, I want these benefits to be small compared to the good sold in the market. Therefore, I will frequently make assumptions that \bar{b} is small, in ways that I make precise below.

¹²That is, advisers cannot recommend to purchase *no* product. This is to simplify exposition. In Section 3.5.2 I extend the model to allow advisers to recommend no purchase.

¹³This is by assumption, but in equilibrium no adviser would like to deviate by disclosing more information. This is because, as I prove in Lemma 1, consumers always follow the adviser's recommendation in equilibrium. Therefore, advisers cannot gain by providing more information.

3. FEE-BASED VERSUS COMMISSION-BASED ADVICE

Adviser remuneration. I consider two modes of adviser remuneration. In the first case, advisers are remunerated by the firms through kickbacks. In this case, advice is free for consumers and an adviser receives a kickback k_i if it sells product i . Kickbacks are set by the firms. Kickbacks are unobserved by consumers, but they form beliefs about them that are correct in equilibrium. Second, I discuss the case where these kickbacks are regulated so that consumers must pay for advice themselves. Every adviser j then posts an advice fee a_j which consumers must pay if they receive advice.¹⁴

Advisers recommend the product $i \in \{H, L\}$ that maximizes¹⁵

$$k_i + \gamma(\varepsilon_i - p_i), \quad (3.1)$$

where $\gamma \in (0, 1)$. Thus, the adviser cares about the kickback it receives from selling a product (under a commission-based model) and the utility the consumer would obtain from the purchase. The concern for suitability γ measures how much weight advisers put on consumers' match value. As shown by Inderst and Ottaviani (2012a), such a concern for suitability arises if an adviser cares directly about the consumer's payoff, he is liable for bad advice and at risk of receiving a fine or because of reputational concerns. For the purpose of calculating welfare, I assume that the second term in (3.1) denotes a fine the adviser must pay if he recommends the wrong product, (the probability of) which is increasing in the amount of misselling.

Differences with Inderst and Ottaviani (2012a). My baseline model for commission-based advice is very similar to Inderst and Ottaviani (2012a), but contains several differences.¹⁶ The first is that Inderst and Ottaviani (2012a) model product differentiation in the vein of Hotelling (1929), while in my model product differentiation is in the form of Perloff and Salop (1985). The main reason to use the Perloff and Salop (1985) approach is that in my model both firms and advisers are remunerated. There can be an arbitrary number of advisers, which is easier to handle using Perloff and Salop (1985). I model

¹⁴Advisers cannot charge an advice fee when they receive kickbacks. This assumption reflects that financial advice, when remunerated through kickbacks, tends to be free. In some cases, this reflects the institutional setting. For example, in The Netherlands, before 2009 it was illegal for financial advisers to charge consumers anything. In other cases, this assumption can be endogenized as a self-confirming equilibrium. To see this, assume firms do not wish to sell through an adviser charging consumers, expecting that no consumer wants to purchase there. Then, it is indeed rational for consumers not to visit such an adviser, making the firms' original inference rational. As a result, a fee-charging adviser makes no profits so that it is an equilibrium that advisers do not charge consumers when they receive kickbacks.

¹⁵The recommendation rule is by assumption for ease of exposition. One could similarly assume, as in Inderst and Ottaviani (2012a), that the consumer and adviser play a cheap talk game in which the adviser's utility is (3.1). It is then always an equilibrium that the advisers recommend the product that maximizes (3.1). Though, there are also babbling equilibria in which no advice is conveyed as well as trivial equilibria where no consumer visits an adviser and the advice given is irrelevant (i.e. any advice is an equilibrium), I focus on the informative equilibrium as derived by Inderst and Ottaviani (2012a).

¹⁶Inderst and Ottaviani (2012a) do not consider fee-based advice. Below I also introduce capacity constraints for advisers, a feature also not present in Inderst and Ottaviani (2012a).

differentiation between firms in the same way to ease exposition. However, it is easy to verify that switching between the two forms of differentiation does not change the results in Inderst and Ottaviani (2012a), so that any differences in results are not caused by this assumption.

A second difference is that in Inderst and Ottaviani (2012a), advisers care only about the match value of the recommended product and not the price. In the notation of my model, advisers then recommend the product that maximizes

$$k_i + \gamma \varepsilon_i$$

instead of (3.1). In my model advisers care also about the affordability of the recommended product. I view this assumption as more natural: for advice to be unbiased does not mean that the adviser recommends the product with just the highest match value, but the product that makes the consumer best off. Because, as I show below, in equilibrium both firms charge the same price, it would seem that changing this assumption is inconsequential.¹⁷ However, off the equilibrium path the difference does matter. Therefore, the efficiency of the resulting equilibrium depends on this assumption. Inderst and Ottaviani (2012a) find that the equilibrium under commission-based advice is always inefficient, I show that this need not be the case. I explain this in more detail below.

3.3 Short-run equilibria

To gain some insight in the model and to aid comparison with Inderst and Ottaviani (2012a), I first analyze the model for a fixed number of advisers. In Section 3.4, I allow for entry and exit.

3.3.1 Short-run equilibrium with kickbacks

First, I consider the case where advisers are remunerated through kickbacks. In this case, the timing of the game is as follows:

1. firms L and H simultaneously set prices p_L , p_H and kickbacks k_L and k_H ;
2. consumers choose which adviser to visit;
3. advisers recommend a product;

¹⁷The equilibrium price of both firms is equal both in Inderst and Ottaviani (2012a) and in my model with commission-based advice.

3. FEE-BASED VERSUS COMMISSION-BASED ADVICE

4. consumers choose which product (if any) to purchase.

I solve the game backwards. First, when do consumers follow advisers' recommendations? To answer this question, it is necessary to make an assumption on the formation of consumer beliefs off the equilibrium path: if a consumer observes a non-equilibrium price, what does he believe about kickbacks? (Recall that consumers observe prices but not kickbacks.) To stay close to Inderst and Ottaviani (2012a), I assume like them that consumers have passive beliefs: they do not change their beliefs on kickbacks when they observe off-equilibrium prices. That is, they believe that firm L has set the equilibrium kickback k_L^* and firm H the equilibrium kickback k_H^* no matter what prices they observe. The following Lemma states that under this assumption, consumers will follow advisers' recommendations both on and off the equilibrium path as long as firms charge prices that are not too high.

Lemma 1. *Consider an advice subgame where firm $i \in \{L, H\}$ has set price p_i and kickback k_i . Assume the consumer has rational but passive beliefs on kickbacks: they believe $k_L = k_L^*$ and $k_H = k_H^*$. The adviser recommends to purchase i . The consumer will follow this recommendation if and only if*

$$p_i \leq \hat{\varepsilon}_i(p_i) \equiv \mathbb{E}[\varepsilon_i | i \text{ recommended}].$$

Otherwise, the consumer does not make a purchase.

To show that consumers follow the adviser's recommendation to purchase i , it is required to show that they prefer purchasing i over $j \neq i$ as well as that they prefer i over the outside option. That they prefer i over j follows from the specification of advisers' utility (3.1). Since advisers are concerned about the suitability of the products they sell, their signals are informative. In particular, passive beliefs imply that the recommendation to purchase i implies that (for $j \neq i$)

$$k_i^* + \gamma(\varepsilon_i - p_i) \geq k_j^* + \gamma(\varepsilon_j - p_j). \quad (3.2)$$

Since the consumer does not know if firm i is the low- or high-cost firm, he thinks that either i is the high-cost firm (with probability $P(\text{adviser recommends } H)$) and

$$k_H^* + \gamma(\varepsilon_i - p_i) \geq k_L^* + \gamma(\varepsilon_j - p_j)$$

or that i is the low-cost firm (with probability $P(\text{adviser recommends } L)$) and

$$k_L^* + \gamma(\varepsilon_i - p_i) \geq k_H^* + \gamma(\varepsilon_j - p_j).$$

Say, for example, that $k_L^* > k_H^*$. Then, if i is the low-cost firm the recommendation does not contain much information since the recommendation is probably caused by

the fact that i pays higher kickbacks. On the other hand, if i is the high-cost firm the recommendation is very informative: the adviser recommended i even though j pays higher kickbacks so the utility difference $(\varepsilon_i - p_i) - (\varepsilon_j - p_j)$ must be sufficiently large. Therefore, a recommendation can only convey positive information. As a result, households prefer purchasing i over j if the recommendation is i .

Second, consumers must prefer to follow the recommendation to purchase i over the outside option. Since consumers' utility of the outside option equals zero, they follow this recommendation if

$$\mathbb{E}[\varepsilon_i | i \text{ recommended}] - p_i \geq 0.$$

This gives the upper bound on prices, $\hat{\varepsilon}_i(p_i)$, in the Lemma. Note that this upper bound depends on the price p_i because the price conveys information on the expected utility of the product: the higher the price, the more utility the consumer must derive from the product in order for the adviser to recommend it.

I now derive equilibrium prices and kickbacks. First, Lemma 1 implies that firm i 's equilibrium price must satisfy $p_i^* \leq \hat{\varepsilon}_i(p_i^*)$: if not, no consumer would ever buy from firm i and it could increase its profits by dropping its price below $\hat{\varepsilon}_i(p_i^*)$. In fact, it must be the case that $p_i^* = \hat{\varepsilon}_i(p_i^*)$. To see this, note that when the equilibrium price is strictly lower than $\hat{\varepsilon}_i(p_i)$, firm i has the following profitable deviation: it can increase its price by ϵ and its kickbacks by $\gamma\epsilon$. This leaves the probability of advisers recommending i constant, but increases i 's margins since $\gamma < 1$. Thus, the equilibrium must satisfy $p_i^* = \hat{\varepsilon}_i(p_i^*)$. Moreover, since consumers cannot differentiate between the high- and low-cost firm, their willingness to pay is the same for both firms. This implies that firm L and firm H charge the same price. Thus $p_H^* = p_L^* = p^* = \hat{\varepsilon}$.¹⁸

The consumer surplus from visiting adviser j is

$$CS_j = \mathbb{E}[\varepsilon | \text{recommended}] - p^* + b_j = \hat{\varepsilon} - p^* + b_j = b_j.$$

Hence, all consumers want to visit an adviser in equilibrium and a consumer visits the adviser that delivers the highest convenience benefits.

Firm $i \in L, H$ chooses its kickbacks to maximize its profits

$$\pi_i^F(k_i) = (p^* - c_i - k_i)P(k_i + \gamma(\varepsilon_i - p^*) \geq k_j + \gamma(\varepsilon_j - p^*)).$$

The first term is the firm's margin, the second term denotes the probability that an adviser recommends product i . The kickback game is equivalent to a game of Bertrand

¹⁸Contrary to $\hat{\varepsilon}_i(p_i)$, $\hat{\varepsilon}$ does not depend on any prices since in equilibrium both firms charge the same price. Therefore prices drop out of equation (3.2) and, in equilibrium, a firm's price is not informative on the match value it provides, conditional on its product being recommended.

3. FEE-BASED VERSUS COMMISSION-BASED ADVICE

competition with differentiated products. To see this, introduce $p'_i \equiv p^* - k_i$. Then firm i 's profits can be written as

$$\pi_i^F(p'_i) = (p'_i - c_i)P\left(\gamma\varepsilon_i - p'_i \geq \gamma\varepsilon_j - p'_j\right). \quad (3.3)$$

This is the profit function of a firm in a game of Bertrand competition where the match values are equal to $\gamma\varepsilon$. It follows from Caplin and Nalebuff (1991) that the game is supermodular and has a unique equilibrium.

Proposition 6. *When advisers are remunerated through kickbacks, there exists a unique equilibrium. The equilibrium price is*

$$p_L^* = p_H^* = p^* = \hat{\varepsilon} \equiv \mathbb{E}[\varepsilon_i | i \text{ recommended}].$$

When firms are asymmetric ($c > 0$), kickbacks satisfy $k_L^ > k_H^*$. All consumers visit an adviser in equilibrium.*

Firm L pays higher kickbacks than firm H . It uses its cost advantage to capture a larger market share not through lower prices, but by providing relatively large inducements to advisers. As a result, advisers steer consumers towards the more efficient firm. However, whether the equilibrium is efficient depends on advisers' concern for suitability γ . By Lemma 1, all consumers purchase in equilibrium, which is efficient since one product has marginal costs of zero. Hence, efficiency requires that the consumer purchases the low-cost product if and only if

$$\varepsilon_L \geq \varepsilon_H - c.$$

Since the adviser recommends L if and only if

$$k_L^* + \gamma(\varepsilon_L - p^*) \geq k_H^* + \gamma(\varepsilon_H - p^*)$$

the equilibrium is efficient if and only if

$$\frac{k_L^* - k_H^*}{\gamma} = c.$$

When the left-hand side is larger, the recommended quantity of L is too high; when the left-hand side is smaller, the recommended quantity of L is too low. Obviously, this condition is satisfied when $c = 0$. Then firms are symmetric and they charge the same kickbacks. When firms are not symmetric, the efficiency of the equilibrium depends on γ . Because kickbacks depend on γ , it is not immediately obvious how $\frac{k_L^* - k_H^*}{\gamma}$ changes with γ . The proof of the following proposition shows that this quantity is strictly decreasing in γ so that there exists a unique cut-off value for γ at which the equilibrium is efficient.

Proposition 7. *If $c > 0$, then if*

$$\frac{k_L^* - k_H^*}{\gamma'} = c.$$

has a solution $\gamma' \in (0, 1)$, it is unique. If $\frac{k_L^ - k_H^*}{\gamma} < c$ for all $\gamma \in (0, 1)$, set $\gamma' = 0$. If $\frac{k_L^* - k_H^*}{\gamma} > c$ for all $\gamma \in (0, 1)$, set $\gamma' = 1$. Efficiency of the equilibrium with kickbacks depends on γ as follows:*

- *the equilibrium is efficient if $c = 0$ or $\gamma = \gamma'$;*
- *the quantity of the low-cost firm is too high if $c > 0$ and $\gamma < \gamma'$;*
- *the quantity of the low-cost firm is too low if $c > 0$ and $\gamma > \gamma'$.*

The observation that the efficiency of the equilibrium depends on the effectiveness of kickbacks in steering advisers' recommendations is contrary to Inderst and Ottaviani (2012a), who find that the market share of the low-cost firm is always too low. This difference occurs, because contrary to my model, in Inderst and Ottaviani (2012a) the adviser only cares about the match value of the recommended product, rather than the overall utility consumers obtain. That is, in their model advisers do not consider the fact that one product might be more expensive than another. While in equilibrium, this difference does not matter—in both their model and mine the firms charge the same price—off the equilibrium path firms might charge different prices and this fact has an effect on the resulting equilibrium. In Inderst and Ottaviani (2012a), a marginal increase in kickbacks increases a firm's market share but has to be paid for inframarginal consumers as well. Therefore, the efficient firm's kickbacks are inefficiently low. The same is true in my model. However, in my model, a marginal increase in kickbacks also allows for an increase in prices, through the double deviation described above.¹⁹ Whether this increase in prices compensates for or exceeds the inframarginal loss caused by higher kickbacks depends on γ . Therefore, advisers' concern for suitability is crucial for efficiency when advisers care about consumers' overall utility rather than just the match value of the recommended product.

I now calculate expected consumer surplus in order to compare it with consumer surplus under direct remuneration below. Because a consumer visits the adviser with the highest convenience benefits, expected consumer surplus is

$$E[CS] = \mathbb{E}[\varepsilon | \text{recommended}] - p^* + \mathbb{E}[\max_j b_j] = \mathbb{E}[\max_j b_j].$$

All surplus from the products is extracted—the only surplus consumers obtain is from the convenience benefits b_j . This extraction is caused by the double deviation of

¹⁹A similar mechanism was found contemporaneously by Teh and Wright (2019).

increasing prices ϵ and kickbacks by $\gamma\epsilon$ described above: this deviation leaves demand unchanged but increases margins. Therefore, firms increase prices until they equal consumers' willingness to pay. In equilibrium, only convenience benefits are left.

3.3.2 Short-run equilibrium without kickbacks

I now derive the equilibrium when kickbacks are banned. Under fee-based remuneration for advice, an equilibrium consists of two prices: the price firms charge for the products and the fee advisers charge for their services. In the absence of kickbacks, advisers always recommend the product that gives consumers the highest utility. Firm i 's profits are then

$$\pi_i^F(p_i) = (p_i - c_i)q_i(p_i, p_j) = (p_i - c_i)P(\epsilon_i - p_i \geq \epsilon_j - p_j). \quad (3.4)$$

The firm pricing game is simply Bertrand competition between two differentiated duopolists. Caplin and Nalebuff (1991) show that the log-concavity of $g(\epsilon)$ is sufficient for this game to have a unique equilibrium. The equilibrium can be characterized by the first order conditions

$$q_i(p_i, p_j) + (p_i - c_i) \frac{\partial q_i}{\partial p_i}(p_i, p_j) = 0 \quad (3.5)$$

for $i \in \{L, H\}$.

Now I turn to the optimal adviser fees. A consumer expects to receive the same advice from every adviser. Therefore a consumer visits adviser i if it offers relatively high convenience benefits. Also, the expected surplus from visiting the adviser must be positive. Denote this expected surplus by

$$\Delta \equiv \mathbb{E}_{\epsilon_L, \epsilon_H} [\max\{\epsilon_L - p_L^*, \epsilon_H - p_H^*\}],$$

where the subscripts on the expectation operators indicate the variables with respect to which the expectation is taken. Taking these two factors into account, adviser i 's profits as a function of its fee a_i and its competitors' fees a_{-i} are

$$\pi_i^A(a_i, a_{-i}) = a_i P\left(b_i - a_i \geq \max_{j \neq i} (b_j - a_j), \Delta + b_i - a_i \geq 0\right) = a_i d_i(a_i, a_{-i}). \quad (3.6)$$

This is again a game of Bertrand competition, but now between the advisers rather than between the firms. In this case, consumers also have the choice not to visit an adviser. The value of this outside option is $-\Delta$. The results of Quint (2014) can be adapted to prove that this game also has a unique equilibrium.²⁰ Because advisers are symmetric,

²⁰The results of Caplin and Nalebuff (1991) are no longer applicable here—they only prove uniqueness for two firms and do not allow for an outside option.

this unique equilibrium is also symmetric and I denote the equilibrium advice fee by a^* . As the proof of the following proposition shows, this advice fee is similarly given by the first order conditions

$$d_i(a^*, a^*) + a^* \frac{\partial d_i}{\partial a_i}(a^*, a^*) = 0. \quad (3.7)$$

Proposition 8. *Under fee-based advice, there exists a unique equilibrium. Firms' prices solve the first order conditions (3.5) and the adviser fee solves the first order condition (3.7).*

Again, the equilibrium is efficient if the low-cost product is recommended when

$$\varepsilon_L \geq \varepsilon_H - c.$$

There now are two sources of inefficiency. The first is that not all consumers might visit an adviser in equilibrium and therefore not all consumers buy. This happens when $\Delta < a^*$, i.e. the expected surplus of the recommended product is smaller than the advice fee. Since the advice fee is increasing in the value of advisers' convenience benefits, this requires that \bar{b} is sufficiently small.²¹ The second source of inefficiency is allocative. Since the low-cost product is recommended when

$$\varepsilon_L - p_L^* \geq \varepsilon_H - p_H^*,$$

the equilibrium is efficient when it is symmetric ($c = 0$) or when

$$p_L^* = p_H^* - c.$$

In other words, when the firms are asymmetric, the equilibrium is efficient when the margin of the high-cost firm, $p_H^* - c$, equals the margin of the low-cost firm, p_L^* . This requires a passthrough of exactly one. As emphasized by Weyl and Fabinger (2013), the amount of passthrough is closely linked to the log-concavity (or log-convexity) of demand. In this case, as the proof of the following proposition shows, log-concavity of demand causes the amount of passthrough to be smaller than one, so that in equilibrium there is insufficient demand for the low-cost product.

Proposition 9. *The equilibrium without kickbacks is only efficient when*

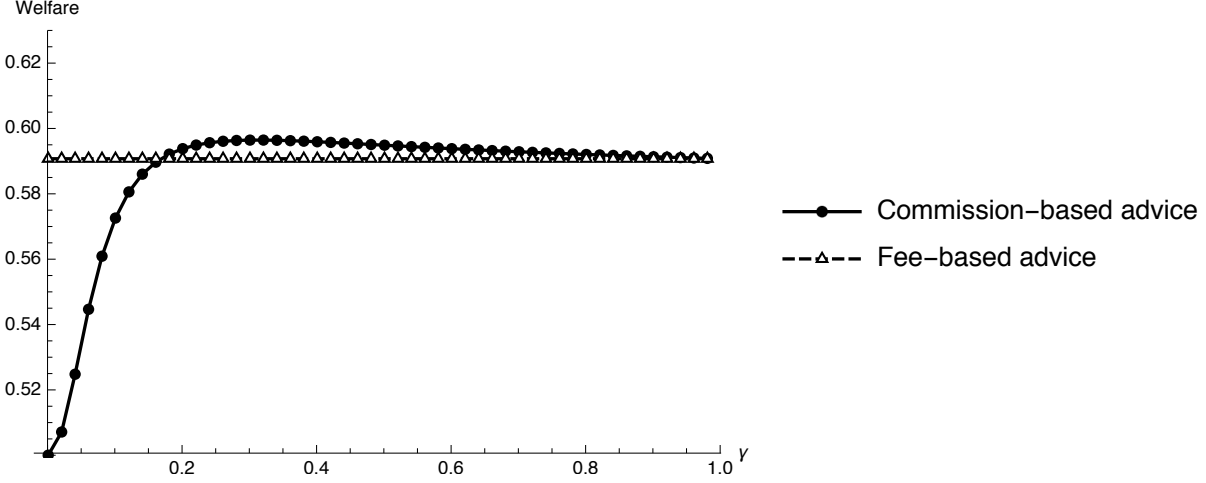
- *all consumers visit an adviser ($\Delta \geq a^*$); and,*
- *the firms are symmetric.*

When $c > 0$, the equilibrium quantity of the low-cost product is inefficiently low.

²¹For example, $\bar{b} \leq \Delta$ is trivially sufficient to guarantee that $a^* \leq \Delta$.

3. FEE-BASED VERSUS COMMISSION-BASED ADVICE

Figure 3.1: Welfare under commission-based and fee-based advice



Note: The figure plots total welfare, excluding advisers' convenience benefits, as a function of advisers' concern for suitability γ . Match values follow distribution $\varepsilon_i \sim Unif(0, 1)$. Firm H has marginal costs $c = \frac{1}{6}$. It is assumed that \bar{b} is sufficiently small, so that all consumers visit an adviser under fee-based advice, i.e. the only efficiency is allocative

Because the equilibrium in the presence of kickbacks is typically also inefficient (Proposition 7), it is not immediately obvious whether the equilibrium is more efficient with or without kickbacks. This question is easy to answer by comparing firms' profit functions without kickbacks (3.4) and firms' transformed profit functions with kickbacks (3.3). As $\gamma \uparrow 1$, the profit functions coincide. Therefore, the allocations are the same with and without kickbacks as γ goes to one. In this case, the equilibria are equally efficient. If γ is between γ' and 1, the equilibrium with kickbacks is more efficient: the closer γ is to the optimal value γ' , the more advisers steer consumers towards the more efficient firm. If γ drops below γ' , advisers steer even more consumers towards the low-cost firm, but now too much so. As γ goes to 0, advisers care less and less about the utility consumers obtain from the recommended product and the probability they recommend the low-cost firm goes to one. Therefore, if the difference in marginal costs c is not too large and γ is sufficiently close to zero, the equilibrium without kickbacks may be more efficient than the equilibrium with kickbacks. Figure 3.1 shows a situation where this happens.

Recall that under kickbacks, consumers only derive surplus from advisers' convenience benefits. Without kickbacks, they do gain positive surplus from the product itself but the convenience benefits are no longer free. Without kickbacks, expected consumer surplus is

$$\mathbb{E}[CS] = \mathbb{E}_{\varepsilon_L, \varepsilon_H}[\max\{\varepsilon_L - p_L^*, \varepsilon_H - p_H^*\}] + \mathbb{E}_b[\max_i b_i] - a^*.$$

Since under kickbacks expected consumer surplus is $\mathbb{E}[\max_i b_i]$, consumers are better

off without kickbacks if and only if

$$\mathbb{E}_{\varepsilon_L, \varepsilon_H}[\max\{\varepsilon_L - p_L^*, \varepsilon_H - p_H^*\}] \geq a^*. \quad (3.8)$$

Since the left-hand side of this equation is the expected surplus from the recommendations consumers receive, this inequality simply states that the benefit of going to an adviser must be greater than the price of advice for a consumer who obtains no convenience benefits. In other words, consumers are better off if they all choose to visit an adviser, i.e. when $\Delta \geq a^*$. This is the case as long as advisers' convenience benefits are not too valuable. Therefore, the following proposition holds.

Proposition 10. *Assume that \bar{b} is sufficiently small, so that, under direct remuneration, all consumers prefer to visit an adviser rather than directly purchase the outside option. Then consumer surplus is higher when advisers are directly remunerated by consumers than when they are remunerated through kickbacks.*

Although the equilibrium is typically more efficient under kickbacks, consumers are—for a fixed number of advisers—always better off under direct remuneration. Because many regulators arguably care more about consumer surplus than total welfare, this finding supports the idea of banning commission-based advice. Holding the market structure fixed, banning kickbacks does increase consumer surplus. However, as I show in the next section, a very different story emerges when accounting for entry and exit of advisers.

3.4 Capacity constraints and entry

So far, I have assumed that one adviser can service the whole market. In reality, this is rarely the case. Giving advice requires learning about the specific circumstances of every consumer. Given that there are only so many working hours in a week, a single adviser can only give advice to so many consumers. Therefore, I now introduce capacity constraints for advisers. Doing so allows me to investigate the claim that banning kickbacks may make advice less accessible. In this section I answer questions like: how does the number of advisers in the market depend on the way they are remunerated? Is entry efficient? How does consumer surplus depend on the number of advisers?

To formalize the idea that a single adviser cannot serve the whole market, I introduce the following capacity constraint.

Assumption 2. *Every adviser can give advice to at most a measure $\lambda = \frac{1}{x}$ of consumers, where x is an integer satisfying $x \geq 2$.*

3. FEE-BASED VERSUS COMMISSION-BASED ADVICE

In principle, all results that follow can be derived for general $\lambda < 1$. However, when there does not exist an integer N such that $\lambda N = 1$, one often has to make case distinctions depending on whether the remainder $\frac{1}{\lambda} \bmod 1$ is large or small.²² To prevent distraction by such irrelevant technicalities, I therefore focus on the case where $\frac{1}{\lambda} \bmod 1 = 0$.

Moreover, I introduce a fixed cost of entry. Advisers can provide consumers with their best match because they have superior expertise. Since this expertise is costly to acquire, an adviser will only enter in markets in which it expects to make sufficient profits. There is a fixed cost of entry $E > 0$ for every adviser, which corresponds to learning about product characteristics or complying with regulations. In the long-run free entry equilibrium, advisers will enter until the profit per adviser net of entry costs E equals zero.²³ To ensure that a pure strategy equilibrium always exists, I assume that advisers' entry is sequential: there are an infinite amount of potential entrants that enter (or not) in a given order. A potential entrant observes the entry decisions of the advisers that move before it. I focus on the case where it is efficient that all consumers receive advice.

Assumption 3. *Entry costs E are small enough that entry of an adviser that recommends all consumers to purchase a random product is efficient:*

$$E < \mathbb{E}[\varepsilon_i].$$

Assumption 3 implies that even if advisers are completely biased towards one product, their entry is still efficient. In other words, I consider markets where it is always efficient that consumers purchase *something*.

As in the rest of this article, I assume that advisers' convenience benefits are small compared to the utility of the primary product. Here, I want to rule out that it can be efficient for advisers to only provide convenience benefits—they have to actually give advice. Therefore, I make the following assumption, which implies that adviser entry is inefficient when an adviser only delivers convenience benefits.

²²The issue is that when $\frac{1}{\lambda} \bmod 1 > 0$, a potential entrant that “covers” the market creates excess capacity as well. Depending on the size of the remainder, the amount of excess capacity can be large or small. When $\frac{1}{\lambda} \bmod 1 = 0$, the entrant that covers the market is always different from the entrant that causes excess capacity.

²³That fixed costs are relevant to entry in the advice market has also been noted by the Financial Conduct Authority in the UK. They say that “providing face-to-face advice has significant fixed costs, which can drive commercial decisions concerning where to focus services” (Financial Conduct Authority 2016, pp. 39).

Assumption 4. Convenience benefits b_i are small enough that entry of an adviser that only offers convenience benefits is inefficient:

$$\mathbb{E}[b_i] < E.$$

3.4.1 Long-run equilibrium with kickbacks

First, I look at entry of advisers when they are remunerated through kickbacks. Observe that the equilibrium in Proposition 6 is independent of the number of advisers N . Therefore, the post-entry equilibrium is the same as with a fixed number of advisers, except that perhaps not all consumers can visit an adviser. Denote by n the measure of consumers visiting an adviser. Since the expected consumer surplus from visiting an adviser is strictly positive, every consumer that can go to an adviser does so. Therefore, $n = \min\{\lambda N, 1\}$.

How many advisers will enter the market? Denote by $\bar{k} = q_L^* k_L^* + q_H^* k_H^*$ the aggregate kickback advisers receive collectively, where q_L^* and q_H^* are the equilibrium quantities sold by firm L and H , respectively. The profit of a single adviser i is then

$$\pi_i^A = \min\left\{\frac{1}{N}, \lambda\right\} \bar{k}.$$

The first term is the quantity of advice adviser i supplies. Because the post-entry equilibrium is symmetric, the demand for every advisor is $\frac{1}{N}$. However, its maximum capacity is λ so that every adviser supplies advice to a measure $\min\left\{\frac{1}{N}, \lambda\right\}$ of consumers.

If $\lambda \bar{k} < E$, no adviser wants to enter the market. If $\lambda \bar{k} \geq E$, at least one adviser wants to enter the market. However, as long as the market is not covered, i.e. $\lambda N < 1$, the profit of every potential entrant adviser is $\lambda \bar{k}$. Thus, if a single adviser wants to enter, advisers enter until the market is covered. Therefore, $\lambda N \geq 1$ and in equilibrium every adviser deals with a fraction $\frac{1}{N} \leq \lambda$ of the market. Under free entry it must then be the case that

$$\pi^A = \frac{\bar{k}}{N} = E.$$

Solving for the equilibrium number of advisers (ignoring the integer constraint) gives the following result.

Proposition 11. Assume that $\lambda \bar{k} \geq E$. Allowing for entry and exit of advisers and when advisers are remunerated through kickbacks, the number of advisers is

$$N_k^* = \frac{\bar{k}}{E}.$$

All consumers visit an adviser ($n^* = 1$).

3. FEE-BASED VERSUS COMMISSION-BASED ADVICE

I now study the efficiency of entry, constraining the post-entry equilibrium to equal that derived in Section 3.3. From Proposition 11, at least $\frac{1}{\lambda}$ advisers enter the market. It is profitable for more advisers to enter when $\frac{\bar{k}}{1+\frac{1}{\lambda}} = \frac{\lambda}{1+\lambda}\bar{k} \geq E$. In this case, there are more advisers than is required to give every consumer advice. Therefore, the entry of the last adviser only affects welfare through delivering more diversity in convenience benefits. Assumption 4 implies that entry that only increases the diversity of convenience benefits is inefficient. When E is slightly larger (but still small enough to induce some entry), exactly sufficient advisers enter the market to help everyone. In this case, Assumption 3 implies that entry is efficient.

Proposition 12. *When advisers are remunerated through kickbacks, there is excessive entry when*

$$\frac{\lambda}{1+\lambda}\bar{k} \geq E.$$

When $E \in \left(\frac{\lambda}{1+\lambda}\bar{k}, \lambda\bar{k}\right]$, entry is efficient.

Intuitively, there is too much entry because of a standard business-stealing externality. A potential entrant does not take into account that some of its customers are already served by existing advisers. Since the potential entrant does make a profit from those consumers, the profit from entry is larger than the welfare gain. Therefore, at the margin entry is profitable, but not welfare-increasing.

3.4.2 Long-run equilibrium without kickbacks

Now I consider entry when advisers are remunerated directly by consumers rather than through kickbacks. Firms' prices do not depend on the number of advisers and are therefore the same as in the short run. However, the fee advisers charge does depend on their number.

As before, I want to focus on the case where convenience benefits are small. In addition to Assumption 4, I therefore make the following assumption.

Assumption 5. *The unique solution to*

$$a = \frac{1 - H(a)}{h(a)}$$

satisfies $a < \Delta$.

This assumption states that a monopolist that *only* sells convenience benefits (and does not give advice), charges a fee that is smaller than Δ . This is simply a way of saying that the convenience benefits are relatively invaluable compared to the surplus

consumers derive from advisers' recommendations. Note that $\bar{b} < \Delta$ is a sufficient condition for Assumption 5. Also note that due the log-concavity of $h(b)$ a unique solution always exists.

As in Section 3.3, an adviser competes with other advisers as well as with the outside option. Adviser i 's profits are

$$\pi_i^A(a_i, a_{-i}) = a_i \min\{\lambda, d_i(a_i, a_{-i})\},$$

with

$$d_i(a_i, a_{-i}) = P\left(b_i - a_i \geq \max_{j \neq i} (b_j - a_j), \Delta + b_i - a_i \geq 0\right)$$

the demand for adviser i . The difference with (3.6) is the introduction of capacity constraints: no adviser can serve more than λ consumers. Note that if $d_i(\cdot) > \lambda$, $\pi_i^A(\cdot)$ is strictly increasing in a_i . Hence, it is never individually rational for an adviser to set a fee such that $d_i(\cdot) > \lambda$ and consumers are never rationed.

As the following proposition shows, the structure of the equilibrium now depends crucially on whether the market is covered ($\lambda N \geq 1$) or not.

Proposition 13. *When capacity-constrained advisers are remunerated by consumers, there exists a unique equilibrium. The equilibrium can be characterized as follows:*

1. $\lambda N \geq 1$: the adviser fee satisfies $a^* \leq \Delta$ and every consumer visits an adviser ($n^* = 1$).
2. $\lambda N < 1$: the adviser fee satisfies $a^* > \Delta$ and every adviser operates exactly at capacity ($n^* = \lambda N$).

The equilibrium prices p_L^* and p_H^* are the same as in Proposition 8, i.e. as without capacity constraints.

The proof of the proposition is quite long. Therefore, I now briefly sketch the main argument. As before, competition between advisers is equivalent to a game of Bertrand competition between differentiated firms. Quint (2014) has recently proven that such a game has a unique equilibrium. His setup, however, does not allow for capacity constraints. I therefore adapt his proof. A key problem is that an adviser's effective demand $\min\{\lambda, d_i(a_i, a_{-i})\}$ is not everywhere differentiable. Therefore, I use a result due to Jensen (2007) on increasing differences for Lipschitz continuous functions to show that when $d_i(\cdot)$ has increasing differences in a_i and a_{-i} (which has been shown by Quint (2014)), so does $\min\{\lambda, d_i(a_i, a_{-i})\}$. Existence of an equilibrium then follows from standard arguments, e.g. Milgrom and Roberts (1990). Uniqueness follows from proving that if there are two potential equilibria, at most one can be at a point of non-differentiability. In the other one, the first order conditions must thus hold. From here, a

3. FEE-BASED VERSUS COMMISSION-BASED ADVICE

similar argument as in Quint (2014) proves based on these first order conditions that in the first potential equilibrium, the adviser can always increase its profits by increasing its fee.

To characterize the equilibrium, I proceed as follows. To prove that when $\lambda N \geq 1$, $a^* < \Delta$, I consider a game where advisers do not provide convenience benefits. Assumption 5 then implies that a monopolist would charge a fee smaller than Δ . Because demand is log-concave, the equilibrium fee is then smaller than Δ for any number of advisers. Then I show that when the equilibrium fee of this related game satisfies $a^* < \Delta$, it is also the equilibrium fee of the original game. The intuition for this result is as follows. An adviser in essence sells a bundle of two goods: the recommended product and its convenience benefit. The recommended product is a homogeneous good: consumers expect every adviser to give the same recommendation and their willingness to pay does not depend on the recommended product. Therefore, when $\lambda N \geq 1$, any potential profits from advice will be competed away as in Bertrand competition with homogeneous products. Hence, the resulting advice fee is exactly the fee that would arise in the Bertrand equilibrium of a market where the advisers *only* provide convenience benefits.

To characterize the equilibrium for the case $\lambda N < 1$, note that in any candidate equilibrium with $a^* < \Delta$ every adviser can profitably deviate by increasing its fee: since every adviser operates at capacity, such an increase does not cause consumers to visit another adviser. Therefore, advisers can raise their prices until every adviser operates exactly at capacity. Because there is some amount of differentiation between advisers in the form of convenience benefits, the Edgeworth paradox does not occur and an equilibrium in pure strategies exists. Therefore, it must be the case that $a^* > \Delta$. I then show that the only symmetric fee which satisfies this in equilibrium is the one where advisers operate exactly at capacity. Here Assumption 5 is crucial. This assumption makes sure that advisers do not have too much market power: if \bar{b} is too large, advisers would rather focus on selling their convenience benefits to a small group of consumers that derive particularly large benefits from visiting an adviser, instead of operating at capacity for a lower fee.

Which type of equilibrium occurs when advisers' entry is endogenous? In a potential equilibrium satisfying $\lambda N \geq 1$, the last adviser to enter only makes profits from selling convenience benefits. Therefore, entry is only profitable when convenience benefits are sufficiently valuable. Assumption 4 rules this case out: even if this adviser would capture all surplus from its convenience benefits (and it does not), entry is not profitable. Therefore the following proposition holds.

Proposition 14. *When advisers are directly remunerated by consumers, the market is not covered: $\lambda N_d^* < 1$.*

Comparing with Proposition 11 immediately gives the following result.

Corollary 1. *More advisers enter when they are remunerated through kickbacks than when they are paid directly by consumers: $N_k^* > N_d^*$*

Recall that entry of advisers is typically excessive when they are remunerated through kickbacks. When they are paid directly by consumers, the opposite result occurs.

Proposition 15. *When capacity-constrained advisers are remunerated by consumers, there is insufficient entry.*

The intuition for this result as follows. Since all advisers operate at capacity, an entrant does not steal customers from existing advisers. Therefore, there is no business-stealing externality. Since, in addition, a potential entrant does not capture the full surplus of its entry, there is insufficient entry.

3.4.3 Consumer surplus with and without kickbacks

For a fixed number of advisers, consumers are better off under fee-based advice than under commission-based advice (Proposition 10). When allowing for entry and exit, the number of advisers is smaller, however, under direct remuneration than under kickbacks. Therefore, advice is more accessible under kickbacks and the ranking from the short run does not carry over to the long run. In the long run, expected consumer surplus under kickbacks is

$$CS_k = \hat{\varepsilon} - p^* + \mathbb{E}[\max_j b_j] = \mathbb{E} \left[\max_{1 \leq j \leq N_k^*} b_j \right].$$

The last equality follows from the fact that the equilibrium price equals the expected surplus of the recommend product, $\hat{\varepsilon}$ (Proposition 11). Under direct remuneration, expected consumer surplus is

$$\begin{aligned} CS_d &= \lambda N_d^* \left(\mathbb{E}[\max\{\varepsilon_L - p_L^*, \varepsilon_H - p_H^*\}] + \mathbb{E} \left[\max_{1 \leq j \leq N_d^*} b_j \mid \max_{1 \leq j \leq N_d^*} b_j + \Delta - a^* \geq 0 \right] - a^* \right) \\ &= \lambda N_d^* \left(\Delta + \mathbb{E} \left[\max_{1 \leq j \leq N_d^*} b_j \mid \max_{1 \leq j \leq N_d^*} b_j + \Delta - a^* \geq 0 \right] - a^* \right). \end{aligned}$$

3. FEE-BASED VERSUS COMMISSION-BASED ADVICE

Only a measure $\lambda N_d^* < 1$ have access to an adviser and can make a purchase. The surplus of those receiving advice consists of three parts. The first is the expected surplus from the recommended product. The second is the expected convenience benefit. The expectation here is truncated because only consumers who can obtain a relatively large convenience benefit visit an adviser. The final term is the advice fee consumers incur.

To compare expected consumer surplus under kickbacks and under direct remuneration, note that

$$P \left(\max_{1 \leq j \leq N_d^*} b_j + \Delta - a^* \geq 0 \right) = \lambda N_d^*.$$

Therefore,

$$\begin{aligned} CS_d &= \mathbb{E} \left[\max_{1 \leq j \leq N_d^*} b_j \mathcal{I} \left(\max_{1 \leq j \leq N_d^*} b_j \geq a^* - \Delta \right) \right] - (a^* - \Delta) \\ &\leq \mathbb{E} \left[\max_{1 \leq j \leq N_d^*} b_j \right] \\ &\leq \mathbb{E} \left[\max_{1 \leq j \leq N_k^*} b_j \right] = CS_k. \end{aligned}$$

Here, $\mathcal{I}(\cdot)$ is the indicator function. The first inequality follows from Proposition 14, which shows that $\lambda N_d^* < 1$ and Proposition 13, which shows that in this case $a^* > \Delta$. The final inequality follows because $N_k^* > N_d^*$ (Corollary 1). This proves the following proposition.

Proposition 16. *Allowing for advisers' entry and exit, consumer surplus is higher when advisers are remunerated through kickbacks than when they are paid directly by consumers.*

The intuition for this result is as follows. Under both remuneration schemes, consumers do not gain any surplus from the primary product. Under kickbacks, this is the case because *firms* set their prices to extract all surplus from the primary product. Under direct remuneration, it is the *advisers* who can set high fees since in the long-run there is insufficient competition to bring these fees down. Therefore, in both cases, the only source of consumer surplus is advisers' convenience benefits. Since there are more advisers under kickbacks than under direct remuneration, the expected convenience benefit is larger under kickbacks. Moreover, advice is free under kickbacks but not under direct remuneration so that consumers are better off when kickbacks are allowed.

Thus, the effect of banning kickbacks on consumer surplus is reversed when accounting for advisors' endogenous entry and exit. For a fixed number of advisors, banning kickbacks always benefits consumers. Accounting for entry and exit however, banning kickbacks never benefits consumers.

The result that consumer surplus decreases after banning kickbacks hinges on the fact that advice is homogeneous. Therefore, a natural question is what would happen if advisers could in some way differentiate themselves. Differentiation is a common response in markets that shift from commission-based to fee-based advice.²⁴ However, while differentiation might increase advisers' profits, it is unlikely to do much for consumers. Horizontal differentiation is compatible with my framework: as long as the differentiation of advisers is small compared to the differentiation of the primary product, i.e. Assumption 4 holds, the logic of the preceding paragraphs holds and consumers are better off with kickbacks. As an alternative, advisers might choose to vertically differentiate themselves. For example, in a model where finding a consumer's best match takes costly effort (as in Inderst and Ottaviani 2012b), some advisers might exert more effort than others and give better recommendations. Vertical differentiation is only profitable when consumers can observe quality differences between advisers. But if this is the case, a higher quality adviser can increase its fee by exactly the extra surplus it creates compared to other advisers. Therefore, advisers can appropriate all gains from vertical differentiation. Advice might then be more accessible, but all consumer surplus is still extracted through advice fees.

3.5 Extensions

In this section, I discuss three extensions to the main model. First, I consider what happens when consumers are not perfectly rational. Second, I allow advisers to recommend to make no purchase. Third, I allow consumers to search for the best match themselves instead of visiting an advisor. After these extensions, I discuss the role of two main assumptions behind my main model: capacity constraints and adviser differentiation.

3.5.1 Behavioral consumers

In this section, I consider the possibility that consumers have behavioral preferences. There is ample evidence that consumers have systematic biases in household finance.²⁵ Moreover, one of the main concerns of regulators is that consumers undervalue the financial product they (might) buy compared to the one-time fee they have to incur to obtain advice. While one crucial role of advisers might indeed be to help consumers

²⁴For example, the Fidelity report cited in the Introduction (Footnote 7) gives as a main recommendation "tailoring offerings for defined client segments" as a response to increased competitiveness.

²⁵Beshears et al. (2018) contains a good recent overview of the literature on behavioral household finance.

3. FEE-BASED VERSUS COMMISSION-BASED ADVICE

with limited knowledge of the financial landscape, it is indeed those customers who might be least likely to seek professional advice, let alone pay for it.²⁶

To model this concern in a simple way, I assume that there is a distinction between consumers' willingness to pay and their experienced utility. Their willingness to pay is simply ε_i , as in the main model. However, consumers systematically undervalue purchasing *any* product, so that their experienced utility is $\varepsilon_i + \eta$ with $\eta > 0$.

Because consumers' willingness to pay is the same, the resulting equilibria are the same as in the main text. However, the implications for consumer welfare become even starker. For a fixed number of advisers, essentially nothing changes as all consumers purchase in equilibrium. However, accounting for entry and exit, the long-run consumer surplus analysis tilts even more in favor of kickbacks. The reason is that under commission-based advice, every consumer visits an adviser while under fee-based advice some consumers are excluded from the market. Therefore, long-run consumer surplus can be written as follows:

$$CS_k = \eta + \mathbb{E} \left[\max_{1 \leq j \leq N_k^*} b_j \right],$$

$$CS_d = n_d^* \left(\eta + \Delta + \mathbb{E} \left[\max_{1 \leq j \leq N_d^*} b_j \mid \max_{1 \leq j \leq N_d^*} b_j \geq a^* - \Delta \right] - a^* \right).$$

When $\eta = 0$, the difference in surplus between the two regimes as before. However, because under fee-based remuneration not all consumers visit an adviser ($N_d^* < 1$), the difference in consumer surplus increases the more consumers undervalue the value of advice.

This result has two implications. First, it means that, perhaps contrary to common intuition, behavioral biases on the consumer side are a reason not to regulate adviser remuneration. Second, it significantly strengthens the result obtained under rational consumers. One potential objection to the result in Section 3.4.3 is that advisers' convenience benefits are assumed to be small. Since, for rational consumers, the long-run difference in consumer surplus is only due to differences in the supplied convenience benefits, this means that the long-run difference in consumer surplus is also small. However, when consumers undervalue the product, there is scope for significant damage from banning kickbacks.

²⁶For example, the UK Financial Conduct Authority notes that “almost half (44%) [of consumers] say they are confused as to how financial services firms can help them manage their financial affairs” (Financial Conduct Authority 2016, p. 23). In the UK market, consumers that visit an adviser have *higher* self-reported knowledge about financial matters than those that do not (Financial Conduct Authority 2017, Table 3.7).

3.5.2 Advisers can recommend to make no purchase

So far, advisers were assumed to always recommend one of the products, even if the recommended product is worse than consumers' outside option. This means that even when kickbacks are banned, advisers are still biased: they recommend some consumers to purchase a product that gives them a worse pay-off than the outside option. In this section, I relax the assumption that advisers must recommend either product.

In particular, in addition to recommending product L or H , an adviser can also recommend to make no purchase. I assume that the utility an adviser derives from recommending the outside option is zero, so that it makes this recommendation if

$$\max\{k_L + \gamma(\varepsilon_L - p_L), k_H + \gamma(\varepsilon_H - p_H)\} \leq 0.$$

Since this model is very similar to the main model, I give a sketch of the derivations of equilibria and focus on the resulting differences.

First, I consider the situation where advisers are remunerated through kickbacks. As in the main model, it is always optimal for consumers to follow advisers' recommendations.²⁷

Lemma 2. *Consider an advice subgame where firm $i \in \{L, H\}$ has set price p_i and kickback k_i . Assume the consumer has rational but passive beliefs on kickbacks: they believe $k_L = k_L^*$ and $k_H = k_H^*$. The adviser recommends to purchase i . The consumer will follow this recommendation if and only if*

$$p_i \leq \hat{\varepsilon}_i(p_i) \equiv \mathbb{E}[\varepsilon_i | i \text{ recommended}].$$

Otherwise, the consumer doesn't make a purchase. Moreover, the consumer always follows the recommendation to make no purchase.

Firms can increase their prices by ϵ and their kickbacks $\gamma\epsilon$, increasing their margins without lowering demand. By the same logic as in the main model, therefore, both firms charge the same equilibrium price which equals consumers' willingness to pay: $p^* = \hat{\varepsilon}(p^*)$. The demand of firm i as a function of its kickbacks is then

$$\pi_i^F(k_i) = (p^* - c_i - k_i)P\left(\underbrace{k_i + \gamma(\varepsilon_i - p^*) \geq \max_{j \neq i} k_j + \gamma(\varepsilon_j - p^*)}_{i \text{ is recommended over } j}, \underbrace{k_i + \gamma(\varepsilon_i - p^*) \geq 0}_{i \text{ is recommended over no purchase}}\right).$$

²⁷Compared to Lemma 1, I also have to prove that consumers follow the recommendation to purchase the outside option.

3. FEE-BASED VERSUS COMMISSION-BASED ADVICE

By again introducing the pseudo-price $p'_i \equiv p^* - k_i$, firm i 's profits can be rewritten as

$$\pi_i^F(p'_i) = (p'_i - c_i)P\left(\gamma\varepsilon_i - p'_i \geq \gamma\varepsilon_j - p'_j, \gamma\varepsilon_i - p'_i \geq -(1-\gamma)p^*\right).$$

This is again a game of Bertrand competition with differentiated products. The only somewhat non-standard aspect is that the “outside option” gives utility $-(1-\gamma)p^*$ instead of zero. This is a special case of a more general game analyzed by Quint (2014).^{28,29} His results imply the existence of a unique equilibrium.

Proposition 17. *When advisers are remunerated through kickbacks and advisers can recommend the outside option, there exists a unique equilibrium. The equilibrium price is*

$$p_L^* = p_H^* = p^* = \hat{\varepsilon} = \mathbb{E}[\varepsilon_i | i \text{ recommended}].$$

When firms are asymmetric ($c > 0$), kickbacks satisfy $k_L^* > k_H^*$.

The introduction of the possibility to recommend the outside option elucidates that advisers in fact have two types of bias. The first, which I call product bias, means that the adviser does not always recommend the product giving the consumer the highest utility. The second, quantity bias, means that customers are recommended a product giving them negative utility. Since an adviser recommends to purchase a product if

$$\max\{k_L + \gamma(\varepsilon_L - p_L), k_H + \gamma(\varepsilon_H - p_H)\} \geq 0 \quad (3.9)$$

there is quantity bias as long as kickbacks are strictly positive. The main model also features quantity bias. However, there it is by assumption because advisers are restricted to recommending one of the products. Here, I show that quantity bias in fact can arise endogenously.

For fee-based advice, the profit function of firm $i \in \{L, H\}$ is

$$\pi_i^F(p_i) = (p_i - c_i)P\left(\underbrace{\varepsilon_i - p_i \geq \varepsilon_j - p_j}_{i \text{ is recommended over } j}, \underbrace{\varepsilon_i - p_i \geq 0}_{i \text{ is recommended over no purchase}}\right).$$

Just as in the model with kickbacks, the results of Quint (2014) imply that there exists a unique equilibrium for the pricing game between firms. Because the objective of the advisers does not change, they set the same fees as before. Under direct remuneration, there is no quantity bias: an adviser only recommends to purchase a product if it is better than the outside option.

²⁸Quint (2014) considers a game where goods can be both complements and substitutes, but it easy to “turn off” the complement part.

²⁹The results of Caplin and Nalebuff (1991) used in the main model no longer apply here, since they consider only a model without outside option.

I now consider the effects of allowing advisers to recommend the outside option on efficiency. Since one of the products has marginal costs equal to zero, efficiency requires all consumers to make a purchase. As in the main model, efficiency moreover requires that L is purchased if and only if

$$\varepsilon_L \geq \varepsilon_H - c.$$

In the case of kickbacks, this means that efficiency requires

$$\begin{aligned} k_L^* &= k_H^* - c, \\ k_L^* &\geq -\gamma p^*. \end{aligned}$$

The first condition ensures, as before, that the quantity of the low-cost product is efficient. The second condition ensures that all consumers purchase.

The introduction of the outside option makes it difficult to give general results on the efficiency of the equilibrium. The reason is that the equilibrium price p^* equals consumers' willingness to pay, which is itself a function of γ : consumers are willing to pay more if γ is higher. Therefore, the right-hand side of $k_L^* \geq -\gamma p^*$ is decreasing in γ . It can be shown that kickbacks are also decreasing in γ , but it is not clear which side of the inequality decreases faster. For this reason, the inverted U-shape of welfare in γ that obtains when advisers cannot recommend the outside option need not carry over to the current model. When match values are uniform, i.e. when $\varepsilon_i \sim \text{Unif}(0, \bar{\varepsilon})$, it can be shown that the first efficiency condition above reduces to

$$\gamma = \frac{\bar{\varepsilon} - c}{3\bar{\varepsilon} - 2c}.$$

When firms are moreover symmetric, i.e. $c = 0$, the second efficiency condition reduces to³⁰

$$\gamma \leq \frac{4}{7}.$$

When $c = 0$, the first condition implies the second condition, so that for $\varepsilon_i \sim \text{Unif}(0, \bar{\varepsilon})$ and $c = 0$ the equilibrium is efficient if and only if $\gamma = \frac{1}{3}$.

Under direct remuneration, advisers recommend to make no purchase with positive probability so that the equilibrium is inefficient. As γ approaches one, firms' objective functions again coincide, so that for γ close to one the equilibria under direct remuneration and under kickbacks are approximately equally efficient. For lower values of γ , there are two effects. First, just as before, kickbacks change the market share of the low-cost firm. Compared with the situation in which advisors cannot recommend the

³⁰When $c > 0$, this condition becomes intractable.

outside option, an additional advantage of kickbacks is that they expand the market. Because of kickbacks, some consumers who obtain negative utility from both products nevertheless make a purchase. Quantity bias, although bad for consumers, increases efficiency by counteracting firms' market power. Market power allows firms to charge positive prices, which without quantity bias would lead to inefficiently low quantities. However, firms can at the same time expand their market shares through higher kickbacks, eliminating the quantity-reducing effect of higher prices.

Thus, for a given value of γ and a fixed number of advisers, allowing advisers to recommend the outside option makes kickbacks more desirable from the standpoint of efficiency, while banning kickbacks becomes even more desirable from the standpoint of consumer welfare.

The free-entry results in Section 3.4 do however not qualitatively change when allowing advisers to recommend the outside option. The reason is that Proposition 16 does not depend on the exact post-entry equilibrium. As long as expected kickbacks are sufficiently large to induce entry of one adviser, the result that consumers are better off with than without kickbacks holds.

3.5.3 Consumers do not require an adviser to make a purchase

In the analysis up to now, consumers had to visit an adviser to purchase the primary product. However, in many markets with advice, consumers can also purchase directly from the upstream suppliers. Given that consumers in the long run receive no surplus from the primary product under either kickbacks or direct remuneration, it is possible that they prefer to purchase without an adviser. Therefore, in this section I allow consumers to also make a purchase directly from one of the firms.

In particular, consumers can also learn their match values using costly search. For this search channel, I employ the model of Choi, Dai, and Kim (2018). In their model, consumers observe the prices of both firms and then perform directed search across firms. They incur a search cost s per visited firm and there is free recall. I embed their model into mine as follows. Consumers choose, after observing prices and the convenience benefits offered by the advisers, whether to search themselves or visit an adviser. For those that visit an adviser, the analysis is as before. To make the comparison between the search channel and the advice channel consistent, the adviser can recommend to make no purchase as in Section 3.5.2. Consumers that choose to search can visit any number of firms (including zero) and either purchase the best option from those they have visited or the outside option. Firms cannot price discriminate

between the search and the advice channel.³¹ To simplify notation in this section, I focus on the case in which firms are symmetric.

Assumption 6. *The firms are symmetric: $c = 0$.*

Consumers search as follows. Define the reservation value ε^* as the solution to

$$s = \int_{\varepsilon^*}^{\infty} (1 - F(\varepsilon)) d\varepsilon.$$

Choi, Dai, and Kim (2018, Proposition 1) show based on Weitzman (1979) that

- consumers visit sellers in descending order of $\varepsilon^* - p_i$,
- consumers stop searching as soon as the set of visited firms \mathcal{N} satisfies $\max\{0, \max_{i \in \mathcal{N}} \varepsilon_i - p_i\} > \max_{j \notin \mathcal{N}} \{\varepsilon^* - p_j\}$.

After they stop searching, consumers purchase the best option from those they have visited, or the outside option.

A consumer chooses the channel giving him the highest expected utility. Corollary 1 of Choi, Dai, and Kim (2018) implies that the expected consumer surplus of search is

$$\mathbb{E} \left[\max \left\{ 0, \max_{i \in \{L, H\}} \{\min\{\varepsilon_i, \varepsilon^*\} - p_i\} \right\} \right].$$

The expected consumer surplus from the purchase the adviser recommends is³²

$$\mathbb{E} \left[\sum_{i \in \{L, H\}} (\varepsilon_i - p_i) \mathcal{I}(k^* + \gamma(\varepsilon_i - p_i) \geq \max\{0, k^* + \gamma(\varepsilon_j - p_j)\}) \right],$$

with $\mathcal{I}(\cdot)$ the indicator function. Denote by Δ the difference in expected consumer surplus between purchasing the adviser's recommended product and searching. If a^* is the equilibrium fee advisers charge (with $a^* = 0$ when kickbacks are allowed), the probability a consumer visits an adviser rather than searches himself is

$$P(\Delta + \max_i b_i \geq a^*).$$

³¹For empirical evidence that prices are typically the same in the search and the advice channel, see Edelman and Wright 2015a. One explanation of this phenomenon is given by Shen and Wright (Forthcoming). They show that if consumers can purchase directly from a firm after receiving advice ("showrooming"), firms find it optimal not to discount direct purchases because this lowers the effectiveness of kickbacks. A similar mechanism would apply to this model.

³²Recall that consumers have passive beliefs. Therefore, they always expect both firms to charge equilibrium kickbacks, even when observing out-of-equilibrium prices. Since I focus on the symmetric case $c = 0$, the equilibrium will be symmetric, with both firms setting the same kickback k^* in equilibrium.

3. FEE-BASED VERSUS COMMISSION-BASED ADVICE

As long as search costs are not too high, searching provides a viable alternative to visiting an adviser. This reduces the market power of advisers under fee-based advice. The expected gain from visiting an adviser, Δ , is smaller, so that for a given a^* fewer consumers visit an adviser. Indeed, the advice fee derived in Section 3.4 increases one-to-one with Δ . This suggests that, as long as the search channel is sufficiently viable, consumers might be better off after banning kickbacks. The following proposition shows that this is indeed the case.³³

Proposition 18. *Assume that an equilibrium exists and that search costs s and advisers' convenience benefits \bar{b} are sufficiently small. Then, allowing for free entry, consumer surplus is higher when advisers are directly remunerated by consumers than when they are remunerated through kickbacks.*

Thus, the desirability of banning kickbacks depends greatly on the market structure. Perhaps counter-intuitively, banning kickbacks is desirable only in markets in which consumers face relatively low search costs. Moreover, this result implies that policies aimed at reducing the so-called “advice gap”—the observation that many households do not use financial advice—may be counterproductive. For example, making advice obligatory, a policy introduced for example in the Dutch mortgage market, removes consumers' outside option. My results predict that the only effect of such a policy is an increase in the fees advisers charge, so that all consumers are worse off.

Because advisers display no product bias in equilibrium for either remuneration mode when firms are symmetric, the assumption that firms are symmetric biases the analysis of consumer surplus towards commission-based advice. Therefore, since Proposition 18 search shows that banning kickbacks can increase consumer surplus when firms are symmetric, this is even more true when firms are asymmetric.

3.5.4 Discussion of assumptions

The main model makes two substantive assumptions to study the effects of endogenous entry in a model in the vein of Inderst and Ottaviani (2012a). The first is that advisers are capacity constrained; the second is that they are horizontally differentiated (but not too much). I now discuss the robustness of my results to these assumptions.

When both capacity constraints and horizontal differentiation are removed from the model, in the absence of kickbacks advisers engage in standard Bertrand competition.

³³One potential issue is that Choi, Dai, and Kim (2018) require the variance of b_i for demand to be log-concave and hence for an equilibrium to exist. Therefore, it may seem that the proposition's requirement that \bar{b} is sufficiently small is contradictory with the existence of an equilibrium. The issue in their paper is that the variance of b_i must be sufficiently large *compared* to the search costs. Since search costs and \bar{b} are both assumed to be small in the proposition, there is not necessarily a contradiction here.

As a result, the equilibrium advice fee equals zero. It is immediate that no adviser enters the market under fee-based remuneration. As a result, consumers are in the long run better off under kickbacks. Hence, my results are robust to these assumptions in the sense that if they both are removed at the same time my main result still obtains, albeit somewhat trivially.

A more complicated picture emerges when I relax one substantive assumption, but not the other. In Appendix 3.B I show that when advisers are not capacity constrained, consumers are better off without commissions as long as advisers' convenience benefits are sufficiently small. Hence, capacity constraints are therefore crucial for my results.

In Appendix 3.B I also discuss the case where advisers are homogeneous, i.e. when they do not offer convenience benefits. Here, the assumption that $1 \bmod \lambda = 0$ (Assumption 2), is not innocuous: in this case, the number of entrants and consumers' welfare depends on the remainder $1 \bmod \lambda$.³⁴ However, this remainder does not have a clear economic interpretation. Therefore, I view differentiation between advisers as crucial to obtain an economically meaningful model.

3.6 Conclusion

This paper has introduced a theoretical model to understand when fee-based advice is preferable to commission-based advice. In the model, advisers recommend one of two horizontally differentiated products to uninformed consumers. Advisers can be paid through kickbacks or directly by consumers. When advisers are remunerated through kickbacks, they can be biased but advice is free. The main contribution of this paper is to study the endogenous market structure under both remuneration schemes. In particular, it considers entry and exit of advisers.

For a fixed number of advisers, the equilibrium under kickbacks is typically more efficient than under direct remuneration. The reason, as in Inderst and Ottaviani (2012a), is that kickbacks help steer consumers to the more efficient firm. However, consumer surplus is lower under kickbacks because advisers sometimes recommend the product with the lower utility.

Allowing for entry and exit of advisers, however, overturns the result that consumers are better off under direct remuneration. Under kickbacks, there is excessive entry because of a business-stealing externality. Under direct remuneration, however, no such externality occurs because all advisers operate exactly at capacity. Therefore, there is insufficient entry under direct remuneration and not all consumers have access to

³⁴When the remainder is close enough to zero, consumers are better off without kickbacks, when it is close enough to one without.

advice. Moreover, this gives significant market power to advisers. As a result, they extract most surplus from consumers. Therefore, allowing for advisers' entry and exit, consumers surplus is lower when kickbacks are banned than when they are not. When consumers can also search for the best match themselves and search costs are not too high, banning kickbacks however benefits them.

The model can explain empirical patterns observed in markets that shift from commission-based to fee-based advice. In particular, it can explain why after a ban on commissions, advisers exit the market due to competitive pressure but the price of advice remains high. The reason is that the market tips from one type of equilibrium to another. A small number of advisers have significant market power, but as soon as there is sufficient capacity to provide advice to all consumers fees are competed towards zero. In the long run, only an equilibrium with a small number of advisers can therefore be sustained, in which the fees extract the consumer surplus from advice.

More generally, this paper has shown the importance of considering the endogenous structure of the market when regulating advice. Regulations to improve the quality of advice can hurt the profitability of advisers. If this leads to exit of advisers and/or higher fees, consumers might ultimately be worse off.

3.A Proofs

Proof of Lemma 1

First, note that since beliefs on kickbacks are passive, consumers believe that a recommendation for i implies that (for $j \neq i$)

$$k_i^* + \gamma(\varepsilon_i - p_i) \geq k_j^* + \gamma(\varepsilon_j - p_j^*).$$

Consumers do not know which firm has high costs and which has low costs.³⁵ Therefore, the expected value of purchasing from i conditional on it being recommended is

$$\begin{aligned} \mathbb{E}[\varepsilon_i - p_i | i \text{ recommended}] = & \rho_L \mathbb{E} \left[\varepsilon_i - p_i \middle| \varepsilon_i - p_i \geq \varepsilon_j - p_j + \frac{k_H^* - k_L^*}{\gamma} \right] \\ & + \rho_H \mathbb{E} \left[\varepsilon_i - p_i \middle| \varepsilon_i - p_i \geq \varepsilon_j - p_j + \frac{k_L^* - k_H^*}{\gamma} \right], \end{aligned}$$

³⁵And, as Proposition 6 shows, they cannot infer this from prices since in equilibrium the high and low cost firm charge the same price.

where ρ_H and ρ_L are the probability that the adviser recommends the high- and low-cost firm, respectively. Similarly,

$$\begin{aligned}\mathbb{E}[\varepsilon_j - p_j | i \text{ recommended}] &= \rho_L \mathbb{E} \left[\varepsilon_j - p_j \middle| \varepsilon_i - p_i \geq \varepsilon_j - p_j + \frac{k_H^* - k_L^*}{\gamma} \right] \\ &\quad + \rho_H \mathbb{E} \left[\varepsilon_j - p_j \middle| \varepsilon_i - p_i \geq \varepsilon_j - p_j + \frac{k_L^* - k_H^*}{\gamma} \right].\end{aligned}$$

After receiving the recommendation to purchase i , the consumer prefers purchasing i rather than j if

$$\mathbb{E}[\varepsilon_i - p_i | i \text{ recommended}] \geq \mathbb{E}[\varepsilon_j - p_j | i \text{ recommended}].$$

Say $k_L^* \geq k_H^*$. Substituting the expressions derived above and rearranging gives that this is equivalent to

$$\begin{aligned}\rho_L \mathbb{E} \left[(\varepsilon_i - p_i) - (\varepsilon_j - p_j) \middle| \varepsilon_i - p_i \geq \varepsilon_j - p_j + \frac{k_H^* - k_L^*}{\gamma} \right] \\ \geq \rho_H \mathbb{E} \left[(\varepsilon_j - p_j) - (\varepsilon_i - p_i) \middle| \varepsilon_i - p_i \geq \varepsilon_j - p_j + \frac{k_L^* - k_H^*}{\gamma} \right]\end{aligned}$$

or

$$\begin{aligned}\rho_L \mathbb{E} \left[(\varepsilon_i - p_i) - (\varepsilon_j - p_j) \middle| (\varepsilon_i - p_i) - (\varepsilon_j - p_j) \geq \frac{k_H^* - k_L^*}{\gamma} \right] \\ \geq \rho_H \mathbb{E} \left[(\varepsilon_j - p_j) - (\varepsilon_i - p_i) \middle| (\varepsilon_j - p_j) - (\varepsilon_i - p_i) \leq \frac{k_H^* - k_L^*}{\gamma} \right]. \quad (3.10)\end{aligned}$$

First assume that $\rho_L \geq \rho_H$. In this case, the left-hand side is greater than $\rho_L \frac{k_H^* - k_L^*}{\gamma}$, the right-hand side smaller. Therefore, this inequality holds and the consumer prefers to purchase i rather than j .

Now assume that $\rho_L < \rho_H$. This means that

$$P \left(\varepsilon_L - \varepsilon_H \geq p_H - p_L + \frac{k_L^* - k_H^*}{\gamma} \right) < \frac{1}{2}.$$

Because $\varepsilon_L - \varepsilon_H$ has a symmetric distribution, this implies that

$$p_L - p_H > \frac{k_L^* - k_H^*}{\gamma} \geq 0.$$

Write $\Delta p = p_L - p_H$ and $\Delta k = \frac{k_L^* - k_H^*}{\gamma}$ and denote by $x \equiv \varepsilon_L - \varepsilon_H$. In other words, $\rho_L < \rho_H$ is equivalent to $P(x - \Delta p \geq \Delta k) < \frac{1}{2}$ and $\Delta p > \Delta k \geq 0$. (3.10) is then equivalent to

$$\int_{\Delta p - \Delta k}^{\bar{\varepsilon}} (x - \Delta p) f(x) dx \geq \int_{-\bar{\varepsilon}}^{\Delta p - \Delta k} (x - \Delta p) f(x) dx, \quad (3.11)$$

3. FEE-BASED VERSUS COMMISSION-BASED ADVICE

where $f(x)$ is the pdf of x . The left-hand side can be written as

$$-(1 - F(\Delta p - \Delta k))\Delta p + \int_{\Delta p - \Delta k}^{\bar{\varepsilon}} x f(x) dx,$$

and the right-hand side as

$$-F(\Delta p - \Delta k)\Delta p + \int_{-\bar{\varepsilon}}^{\Delta p - \Delta k} x f(x) dx,$$

where $F(x)$ is the cdf of x . Note that x is symmetric around 0. Hence,

$$\int_{-\bar{\varepsilon}}^{\Delta p - \Delta k} x f(x) dx = \int_{-\bar{\varepsilon}}^{\Delta k - \Delta p} x f(x) dx < 0,$$

where the inequality follows from $\Delta k < \Delta p$. Combining, (3.11) can be written as

$$\int_{\Delta p - \Delta k}^{\bar{\varepsilon}} x f(x) dx \geq 1 - 2F(\Delta p - \Delta k) + \int_{-\bar{\varepsilon}}^{\Delta k - \Delta p} x f(x) dx.$$

Since $\Delta p > \Delta k$, the left-hand side is positive. Moreover, $\Delta p > \Delta k$ and the symmetry of x around zero imply that $F(\Delta p - \Delta k) > \frac{1}{2}$. Hence, the right-hand side is negative. As a result, (3.11) and therefore also (3.10) hold when $\rho_L < \rho_H$.

If instead $k_H^* \geq k_L^*$, rewrite $\mathbb{E}[\varepsilon_i - p_i | i \text{ recommended}] \geq \mathbb{E}[\varepsilon_j - p_j | i \text{ recommended}]$ as

$$\begin{aligned} \rho_H \mathbb{E} \left[(\varepsilon_i - p_i) - (\varepsilon_j - p_j) \middle| \varepsilon_i - p_i \geq \varepsilon_j - p_j + \frac{k_L^* - k_H^*}{\gamma} \right] \\ \geq \rho_L \mathbb{E} \left[(\varepsilon_j - p_j) - (\varepsilon_i - p_i) \middle| \varepsilon_i - p_i \geq \varepsilon_j - p_j + \frac{k_H^* - k_L^*}{\gamma} \right]. \end{aligned}$$

From here the same steps as above show that the left-hand side is larger than the right-hand side.

The consumer purchases i rather than no product if

$$p_i \leq \mathbb{E}[\varepsilon_i | i \text{ recommended}] \equiv \hat{\varepsilon}(p_i),$$

where $\hat{\varepsilon}(p_i)$ is consumers' willingness to pay for i .³⁶ Thus, the consumer always follows the recommendation to purchase product i as long as $p_i \leq \hat{\varepsilon}(p_i)$. When the price is higher, there is no demand for product i .

³⁶The willingness to pay depends on p_i since a higher price implies the adviser is less likely to recommend i . Therefore, conditional on the adviser recommending i , an increase in p_i signals that ε_i is larger.

Proof of Proposition 6

Using the transformation $p'_i = p^* - k_i$, firm i 's profit is

$$\pi_i^F(p'_i, p'_j) = (p'_i - c_i)q_i(p'_i, p'_j)$$

with

$$q_i(p'_i, p'_j) = P(\gamma\varepsilon_i - p'_i \geq \gamma\varepsilon_j - p'_j) = \int_0^{\gamma\bar{\varepsilon}} G\left(\frac{\varepsilon + p'_j - p'_i}{\gamma}\right) \frac{g(\varepsilon)}{\gamma} d\varepsilon$$

the demand for product i . Supermodularity of this game, as well as existence and uniqueness of the equilibrium follow directly from Proposition 6 in Caplin and Nalebuff (1991). It also follows from their results that the log-concavity of $g(\varepsilon)$ implies log-concavity of the demand function. As a result, profit functions are quasi-concave and the first order conditions are sufficient for profit maximization. Note that the equilibrium quantities of this game do not depend on p^* .

This argument proves that, for a given p^* , there exist unique kickbacks $k_L^* = p^* - p'_L$ and $k_H^* = p^* - p'_H$ that maximize firms' profits. It remains to be proven that p^* itself is unique. By the reasoning in the main text, it must be the case that

$$p^* = \hat{\varepsilon}(p^*, k_L^*, k_H^*) = \hat{\varepsilon}(p^*, p^* - p'_L, p^* - p'_H).$$

The right-hand side can be written as

$$\begin{aligned} \hat{\varepsilon}(p^*, p^* - p'_L, p^* - p'_H) &= \rho_L \mathbb{E} \left[\varepsilon_L \left| \varepsilon_L - p^* + \frac{p^* - p'_L}{\gamma} \geq \varepsilon_H - p^* + \frac{p^* - p'_H}{\gamma} \right. \right] \\ &\quad + \rho_H \mathbb{E} \left[\varepsilon_H \left| \varepsilon_H - p^* + \frac{p^* - p'_H}{\gamma} \geq \varepsilon_L - p^* + \frac{p^* - p'_L}{\gamma} \right. \right], \end{aligned}$$

where ρ_L and ρ_H denote the probabilities that the adviser recommends L and H , respectively. Simplifying gives that

$$\hat{\varepsilon}(p^*, p^* - p'_L, p^* - p'_H) = \rho_L \mathbb{E} \left[\varepsilon_L \left| \varepsilon_L - \frac{p'_L}{\gamma} \geq \varepsilon_H - \frac{p'_H}{\gamma} \right. \right] + \rho_H \mathbb{E} \left[\varepsilon_H \left| \varepsilon_H - \frac{p'_H}{\gamma} \geq \varepsilon_L - \frac{p'_L}{\gamma} \right. \right].$$

Since $\rho_L = P\left(\varepsilon_L - p^* + \frac{p^* - p'_L}{\gamma} \geq \varepsilon_H - p^* + \frac{p^* - p'_H}{\gamma}\right) = P\left(\varepsilon_L - \frac{p'_L}{\gamma} \geq \varepsilon_H - \frac{p'_H}{\gamma}\right)$ and $\rho_H = 1 - \rho_L$, $\hat{\varepsilon}(p^*, p^* - p'_L, p^* - p'_H)$ does not depend on p^* . It follows that $p^* = \hat{\varepsilon}(p^*, p^* - p'_L, p^* - p'_H)$ has a unique (trivial) solution in p^* and that there exists a unique equilibrium.

The first order condition for firm i is

$$q_i(p'_i, p'_j) + (p'_i - c_i) \frac{\partial q_i}{\partial p'_i}(p'_i, p'_j) = 0.$$

3. FEE-BASED VERSUS COMMISSION-BASED ADVICE

To prove that when $c > 0$, $k_L^* > k_H^*$, assume to the contrary that $k_L^* \leq k_H^*$. This is equivalent to assuming $p_L' \geq p_H'$. Now,

$$\begin{aligned} \frac{\partial \pi_L^F}{\partial p_L'}(p_H', p_H') &= q_L(p_H', p_H') + p_H' q_L'(p_H', p_H') \\ &= q_H(p_H', p_H') + p_H' q_H'(p_H', p_H') \\ &\leq q_H(p_H', p_L') + p_H' q_H'(p_H', p_L') \\ &= c q_H'(p_H', p_L') \\ &< 0. \end{aligned}$$

The first inequality follows from the supermodularity of the firms' profit functions and the final equality substitutes H 's first order conditions. Since π_L^F is quasi-concave, this implies that p_H' is to the right of the maximum with respect to p of $\pi_L^F(p, p_H')$. Because $p_L' \geq p_H'$, p_L' is then also to the right of the maximum of $\pi_L^F(p, p_H')$. But then p_L' is not a best response to p_H' . By contradiction, it follows that in equilibrium $p_L' < p_H'$, or $k_L^* > k_H^*$.

Proof of Proposition 7

Observe from firms' transformed profit functions (3.3) that γ can be seen as a simple scaling of the match values. Therefore, if γ doubles, in equilibrium p_i' also doubles. In other words, p_L' and p_H' are linear homogeneous functions of γ . Therefore,

$$\frac{\partial p_i'}{\partial \gamma} = p_i'$$

for $i \in \{L, H\}$. Then,

$$\frac{\partial}{\partial \gamma} \frac{k_L^* - k_H^*}{\gamma} = \frac{\partial}{\partial \gamma} \frac{p_H' - p_L'}{\gamma} = \frac{p_H' - p_L'}{\gamma} - \frac{p_H' - p_L'}{\gamma^2} < 0,$$

since $\gamma \in (0, 1)$ and, from the proof of Proposition 6, $p_H' > p_L'$. Therefore, $\frac{k_L^* - k_H^*}{\gamma}$ is decreasing in γ and there is at most one γ' solving

$$\frac{k_L^* - k_H^*}{\gamma'} = c.$$

The rest of the proposition follows from the discussion in the main text.

Proof of Proposition 8

I leave the proof of existence and uniqueness of the game between advisers to the more general case, when there are capacity constraints. In other words, uniqueness and existence follow from the proof of Proposition 13, taking advisers' capacity $\lambda = 1$. The remaining statements follow the discussion in the main text.

Proof of Proposition 9

I have to prove that when $c > 0$, $p_H^* - c < p_L^*$. Assume to the contrary that $p_H^* - c \geq p_L^*$. Since $\log \pi_i^F(\cdot)$ has the same maximizer as $\pi_i^F(\cdot)$, the first order condition for firm i can also be written as

$$\frac{\partial \log \pi_i^F}{\partial p_i} = \frac{1}{p_i - c_i} + \frac{\partial \log q_i}{\partial p_i} = 0.$$

Observe that

$$\begin{aligned} \frac{\partial \log \pi_L^F}{\partial p_L}(p_H^* - c) &= \frac{1}{p_H^* - c} + \frac{\partial \log q_L}{\partial p_L}(p_H^* - c, p_H^*) \\ &> \frac{1}{p_H^* - c} + \frac{\partial \log q_L}{\partial p_L}(p_H^*, p_H^*) = \frac{1}{p_H^* - c} + \frac{\partial \log q_H}{\partial p_H}(p_H^*, p_H^*) = 0. \end{aligned}$$

The inequality follows from the log-concavity of $q_L(\cdot)$ and the assumptions that $p_H^* - c \geq p_L^*$ and $c > 0$. The final term is just the first order condition for H , which equals zero. Since

$$\frac{\partial^2 \log \pi_i^F}{\partial p_i^2} = -\frac{1}{(p_i - c_i)^2} + \frac{\partial^2 \log q_i}{\partial p_i^2} < 0$$

by the log-concavity of $q_i(\cdot)$, $\log \pi_i^F(\cdot)$ is concave in p_i . Then, $\frac{\partial \log \pi_L^F}{\partial p_L}(p_H^* - c) > 0$ implies that $p_H^* - c$ is to the left of the maximizer of $\log \pi_i^F(\cdot)$. Since, by assumption, $p_H^* - c \geq p_L^*$, p_L^* is to the left of the maximizer of $\log \pi_i^F(\cdot)$ as well. But this contradicts that p_L^* is the equilibrium price of firm L . By contradiction, I conclude that $p_H^* - c < p_L^*$.

Proof of Proposition 12

When $\frac{\lambda}{1+\lambda} \bar{k} \geq E$, at least one more adviser wants to enter than is necessary to cover the market. I prove that the additional welfare caused by this entry is smaller than $\mathbb{E}[b_i]$. Assumption 4 then implies that entry is excessive. To prove this statement, note that

$$\begin{aligned} \mathbb{E} \left[\max_{i=1, \dots, N} b_i \right] - \mathbb{E} \left[\max_{i=1, \dots, N-1} b_i \right] - \mathbb{E}[b] &= \int N b H(b)^{N-1} h(b) db \\ &\quad - \int (N-1) b H(b)^{N-2} h(b) db - \mathbb{E}[b] \\ &= \mathbb{E} \left[N b H(b)^{N-1} \right] - \mathbb{E} \left[(N-1) b H(b)^{N-2} \right] \\ &\quad - \mathbb{E}[b] \\ &< \mathbb{E} \left[N b H(b)^{N-1} \right] - \mathbb{E} \left[N b H(b)^{N-2} \right] \\ &< 0. \end{aligned}$$

The first inequality follows because $H(b)$ is a cdf, so that $H(b)^{N-2} b \leq b$ for all b (with a strict inequality for some b) and $\mathbb{E} [b H(b)^{N-2}] < \mathbb{E}[b]$. The second inequality follows

from the fact that $H(b) \in [0, 1]$ for all b so that $NbH(b)^{N-1} \leq NbH(b)^{N-2}$ for all b (with again a strict inequality for some b).

When $E \in \left(\frac{\lambda}{1+\lambda} \bar{k} \geq E, \lambda \bar{k} \right]$, $\frac{1}{\lambda}$ advisers enter and the market is exactly covered. Since all consumers purchase in equilibrium (Proposition 6), Assumption 3 immediately implies that entry is efficient.

Proof of Proposition 13

The proof uses the following proposition on increasing differences for non-differentiable functions, which is due to Jensen (2007). Although the following preliminaries do not play a role in my proof, I state them to complete the understanding of the stated proposition. If $X = (\mathbb{R}^N, \succeq_X)$ is an ordered vector space, its positive cone is $X_+ = \{x \in \mathbb{R}^N : x \succeq_X 0\}$. The cone X_+ has the matrix representation $\mathbf{Q} \in \mathbb{R}^{N \times P}$ if its columns (p_1, \dots, p_P) satisfy $X_+ = \{x \in X : x = \sum_{p=1}^P \lambda_i p_i, \lambda_i \geq 0, i = 1, \dots, P\}$. In the following $\partial_t f(x, t)$ denotes the subdifferential of $f(\cdot)$ with respect to t .

Proposition 19 (Jensen (2007)). *Let $X = (\mathbb{R}^N, \succeq_X)$ and $T = (\mathbb{R}^M, \succeq_T)$ be ordered vector spaces. Assume that $f : X \times T \rightarrow \mathbb{R}$ is locally Lipschitz continuous in $t \in T$ and that T 's positive cone T_+ is closed with matrix representation $\mathbf{P} \in \mathbb{R}^{M \times P}$. Define*

$$\mathbf{P}^T \partial_t f(x, t) = \{q \in \mathbb{R}^P : q = \mathbf{P}^T v, v \in \partial_t f(x, t)\}$$

whose section at $t \in T$ is a well-defined convex-valued multifunction, $\mathbf{P}^T \partial_t f(\cdot, t) : x \rightarrow 2^{\mathbb{R}^P}$. $f(x, t)$ then has increasing differences in (x, t) if and only if for every $t \in T$ there exists an order-preserving selection $\xi(x) \in \mathbf{P}^T \partial_t f(x, t)$, for all $x \in X$, i.e., $\xi : X \rightarrow \mathbb{R}^P$ such that $\xi(x') - \xi(x) \geq_P 0$ whenever $x' \succeq_X x$.

Existence and uniqueness of the equilibrium

Consider an adviser i . Denote by $d_i(a_i, a_{-i})$ the fraction of consumers who want to visit adviser i . This means that i is preferred over all other advisers who are not at maximum capacity and over visiting no adviser. Here a_i is the fee charged by adviser i and a_{-i} contains the fees charged by its competitors. Since i is capacity constrained, the number of consumers visiting it is $\min\{d_i, \lambda\}$ and its profits are

$$\pi_i^A(a_i, a_{-i}) = \min\{d_i(a_i, a_{-i}), \lambda\} a_i.$$

I adapt the proof of Lemma 1 in Quint (2014) to prove existence and uniqueness of the equilibrium. Because of the log-concavity of $h(b)$, Theorem 1 in Quint (2014) implies that $d_i(a_i, a_{-i})$ is continuous, differentiable and log-concave in a_i and that $\log d_i(a_i, a_{-i})$ has strictly increasing differences in a_i and a_{-i} .

Lemma 1 in Quint (2014) does not allow for capacity constraints. Therefore, I show in the following lemma that increasing differences of $\log d(\cdot)$ imply increasing differences of $\log \min\{d_i(a_i, a_{-i}), \lambda\}$.

Lemma 3. $\tilde{d}(a_i, a_{-i}) \equiv \log \min\{d_i(a_i, a_{-i}), \lambda\}$ has increasing differences in a_i and a_{-i} .

Proof. I use a proposition due to Jensen (2007), which for completeness I have restated as Proposition 19 above.

Take for $X = (\mathbb{R}, \geq)$ and $T = (\mathbb{R}^{N-1}, \geq)$, i.e. X and T are vector spaces equipped with the normal element-wise vector order. The positive cone of T is then $T_+ = \mathbb{R}_+^{N-1}$, which is closed under \geq . It is trivial to see that T_+ is generated by the $(N-1) \times (N-1)$ identity matrix, \mathbf{I}_{N-1} . Denote by $\partial_{a_{-i}} \tilde{d}(a_i, a_{-i})$ the subdifferential of $\tilde{d}(\cdot)$ with respect to a_{-i} .

Note that

$$\tilde{d}(a_i, a_{-i}) = \log \min\{d_i(a_i, a_{-i}), \lambda\} = \min\{\log d_i(a_i, a_{-i}), \log \lambda\}.$$

Since $d_i(\cdot)$ is continuously differentiable, $\tilde{d}(\cdot)$ is Lipschitz continuous. Hence, to prove that $\tilde{d}(\cdot)$ has increasing differences in a_i and a_{-i} , the theorem requires that for every $a_{-i} \in T$ there exists a selection $\xi : X \rightarrow \mathbb{R}^{N-1}$, $\xi(a_i) \in \mathbf{I}_{N-1}^T \partial_{a_{-i}} \tilde{d}(a_i, a_{-i}) = \partial_{a_{-i}} \tilde{d}(a_i, a_{-i})$, such that for all a'_i, a_i such that $a'_i \geq a_i$, $\xi(a'_i) - \xi(a_i) \geq 0$. I now construct such a selection. When $d(a_i, a_{-i}) < \lambda$, the subdifferential contains only one element, so that $\xi(a_i) = \frac{\partial \tilde{d}}{\partial a_{-i}}$. Similarly, when $d(a_i, a_{-i}) > \lambda$, the only choice is $\xi(a_i) = \frac{\partial \tilde{d}}{\partial a_{-i}} = 0$. When $d(a_i, a_{-i}) = \lambda$, I choose $\xi(a_i) = 0$ for all a_i .³⁷

There are three possible cases. First, consider the case that $d_i(a'_i, a_{-i}) \leq d_i(a_i, a_{-i}) < \lambda$. Then

$$\xi(a'_i) - \xi(a_i) = \frac{\partial \tilde{d}_i}{\partial a_{-i}}(a'_i, a_{-i}) - \frac{\partial \tilde{d}_i}{\partial a_{-i}}(a_i, a_{-i}) = \frac{\partial \log d_i}{\partial a_{-i}}(a'_i, a_{-i}) - \frac{\partial \log d_i}{\partial a_{-i}}(a_i, a_{-i}) \geq 0,$$

because $\log d_i(\cdot)$ has increasing differences in a_i and a_{-i} .

The second case is $d_i(a'_i, a_{-i}) < \lambda \leq d_i(a_i, a_{-i})$. In this case,

$$\xi(a'_i) - \xi(a_i) = \frac{\partial \tilde{d}_i}{\partial a_{-i}}(a'_i, a_{-i}) - 0 = \frac{\partial \log d_i}{\partial a_{-i}}(a'_i, a_{-i}) \geq 0,$$

since adviser i 's demand increases in its competitors' fees.

The final case is $\lambda \leq d_i(a'_i, a_{-i}) \leq d_i(a_i, a_{-i})$. In this case, $\xi(a'_i) - \xi(a_i) = 0 - 0 = 0$. Hence, the selection $\xi(\cdot)$ satisfies the requirement that $\xi(a'_i) - \xi(a_i) \geq 0$ and $\tilde{d}(\cdot)$ has increasing differences in a_i and a_{-i} . \square

³⁷The subdifferential of the minimum is the convex hull of the derivatives. Hence, the subdifferential contains 0.

3. FEE-BASED VERSUS COMMISSION-BASED ADVICE

To prove existence, first note that it is never optimal to set an advice fee $a_i = 0$.³⁸ Hence, adviser i 's profit maximization problem can be written as

$$\begin{aligned} & \text{maximize } \bar{u}(a_i, a_{-i}) \equiv \log(a_i) + \tilde{d}_i(a_i, a_{-i}) = \log(a_i) + \log \min\{d_i(a_i, a_{-i}), \lambda\} \\ & \text{subject to } \underline{a} \leq a_i \leq \Delta + \bar{b}, \end{aligned}$$

for some $\underline{a} > 0$. Because the second term of $\bar{u}(\cdot)$ has increasing differences in a_i and a_{-i} by Lemma 3, $\bar{u}(\cdot)$ also has increasing differences in a_i and a_{-i} . Hence, the game between advisers is supermodular on bounded strategy spaces. It follows from standard arguments (Milgrom and Roberts 1990) that an equilibrium exists.

To prove uniqueness of the equilibrium, it is useful to first transform the problem again. As noted in the main text, profit maximization implies that $d_i(\cdot) \leq \lambda$. Hence, adviser's i profit maximization problem can be written as³⁹

$$\begin{aligned} & \text{maximize } u(a_i, a_{-i}) \equiv \log(a_i) + \log(d_i(a_i, a_{-i})) \\ & \text{subject to } d_i^{-1}(\lambda, a_{-i}) \leq a_i \leq \Delta + \bar{b}. \end{aligned}$$

To prove uniqueness, assume there are two equilibria $(\bar{a}_0, \bar{a}_1, \dots, \bar{a}_N)$ and $(\bar{a}'_0, \bar{a}'_1, \dots, \bar{a}'_N)$. In the definition of the equilibrium vector I have also included the “price” of the outside option. Of course, $\bar{a}_0 = \bar{a}'_0 = 0$. Note that when a_0 is included in a_{-i} , $d_i(a_i, a_{-i}) = d_i(a_i - \varepsilon, a_{-i} - \varepsilon)$ for all ε ; decreasing *all* prices, including that of the outside option, does not change demand. Because advisers are symmetric, it is possible without loss of generality to assume $\bar{a}'_1 > \bar{a}_1$ and $\bar{a}'_i - \bar{a}_i \leq \bar{a}'_1 - \bar{a}_1$ for all $i > 1$. Denote $\varepsilon = \bar{a}'_1 - \bar{a}_1$.

First observe that it must be the case that $d_1(\bar{a}') < \lambda$. To see why, assume the contrary. Then,

$$\lambda = d_1(\bar{a}'_1, \bar{a}'_{-1}) = d_1(\bar{a}'_1 - \varepsilon, \bar{a}'_{-1} - \varepsilon) = d_1(\bar{a}_1, \bar{a}'_{-1} - \varepsilon) < d_1(\bar{a}_1, \bar{a}_{-1}).$$

The second equality follows because subtracting ε from all prices (including that of the outside option) does not change demand. The inequality follows since $\bar{a}_i \geq \bar{a}'_i - \varepsilon$ for all $i > 1$, $\bar{a}_0 = 0 > \bar{a}'_0 - \varepsilon = -\varepsilon$, and 1's demand increases in its competitors' fees. But since profit maximization requires that $d_i(\cdot) \leq \lambda$, this contradicts that \bar{a} is an equilibrium.

³⁸As in Quint (2014), this can easily be shown by observing that when all i 's competitors set a fee of zero, i 's best response, call it \underline{a} , is strictly positive. Since the game is supermodular it is then possible to restrict i 's strategy space to $[\underline{a}, \Delta + \bar{b}]$ without loss of generality.

³⁹This formulation cannot be used to prove existence of the equilibrium, since the set $\{a \in [0, \Delta + \bar{b}]^N : d_i(a) \leq \lambda, \forall i\}$ is not a complete lattice.

Since \bar{a}'_1 is in the interior of adviser 1's constraint set, the first order conditions $\frac{\partial u}{\partial a_1}(\bar{a}') = 0$ are necessary.⁴⁰ From these first order conditions it follows that

$$\begin{aligned}
 0 = \frac{\partial u_1}{\partial a_1}(\bar{a}'_1, \bar{a}'_{-1}) &= \frac{1}{\bar{a}'_1} + \frac{\partial d_1}{\partial a_1}(\bar{a}'_1, \bar{a}'_{-1}) \\
 &= \frac{1}{\bar{a}'_1} + \frac{\partial d_1}{\partial a_1}(\bar{a}'_1 - \varepsilon, \bar{a}'_{-1} - \varepsilon) \\
 &= \frac{1}{\bar{a}'_1} + \frac{\partial d_1}{\partial a_1}(\bar{a}_1, \bar{a}'_{-1} - \varepsilon) \\
 &< \frac{1}{\bar{a}'_1} + \frac{\partial d_1}{\partial a_1}(\bar{a}_1, \bar{a}_{-1}) \\
 &< \frac{1}{\bar{a}_1} + \frac{\partial d_1}{\partial a_1}(\bar{a}_1, \bar{a}_{-1}) \\
 &= \frac{\partial u_1}{\partial a_1}(\bar{a}_1, \bar{a}_{-1}).
 \end{aligned}$$

The first inequality follows from the fact that $d(\cdot)$ has strictly increasing differences in a_i and a_{-i} . But this means that \bar{a}_1 is not a best response: firm 1 can increase its profits by increasing its fee. Note that this is always possible by the assumption that $\bar{a}_1 < \bar{a}'_1$. As a result, the equilibrium is unique.

Characterization of the equilibrium

I characterize the equilibrium separately for the case $\lambda N < 1$ and for the case $\lambda N > 1$.

Case 1: $\lambda N > 1$. Because the advisers are symmetric, the unique equilibrium will be symmetric as well. This means that in equilibrium, every adviser serves fewer consumers than its capacity: if every adviser would serve a number of consumers equal to its capacity, the total number of consumers served would be $\lambda N > 1$, i.e. larger than the number of consumers in the market. Therefore, the capacity constraint is never binding in equilibrium and the concavity of $u_i(\cdot)$ implies that the unique equilibrium is characterized by the first order condition

$$\frac{\partial u_i}{\partial a_i}(a^*, a^*) = \frac{1}{a^*} + \frac{\partial \log d_i}{\partial a_i}(a^*, a^*) = 0. \quad (3.12)$$

To prove that the equilibrium fee satisfies $a^* \leq \Delta$, consider for now Δ as an exogenous parameter of the model and denote the equilibrium fee by $a^*(\Delta)$. When $\Delta = 0$, advisers only deliver convenience benefits and do not give advice. Assumption 5 says that in

⁴⁰Note that, contrary to $\bar{u}(\cdot)$, $u(\cdot)$ is differentiable everywhere.

3. FEE-BASED VERSUS COMMISSION-BASED ADVICE

this case, a monopolist would charge a fee smaller than Δ . The log-concavity of $h(b)$ then implies that $a^*(0) < \Delta$ for all N .⁴¹

The next step is to prove that when $a^* < \Delta$, a^* is in fact independent from Δ . To see this, note that the demand of a potentially deviant adviser i is

$$d_i(a_i, a^*) = P(b_i - a_i \geq \max_{j \neq i} b_j - a^*, b_i + \Delta - a_i \geq 0) = \int_{a_i - \Delta}^{\bar{b}} H(b + a^* - a_i)^{N-1} h(b) db.$$

$d_i(\cdot)$ has first and second derivatives

$$\begin{aligned} \frac{\partial d_i}{\partial a_i}(a_i, a^*) &= - (N-1) \int_{a_i - \Delta}^{\bar{b}} H(b + a^* - a_i)^{N-2} h(b + a^* - a_i) h(b) db \\ &\quad - H(a^* - \Delta)^{N-1} h(a_i - \Delta), \\ \frac{\partial^2 d_i}{\partial a_i \partial \Delta}(a_i, a^*) &= - (N-1) H(a^* - \Delta)^{N-2} h(a^* - \Delta) h(a_i - \Delta) \\ &\quad + (N-1) H(a^* - \Delta)^{N-2} h(a^* - \Delta) h(a_i - \Delta) \\ &\quad + H(a^* - \Delta)^{N-1} h'(a_i - \Delta). \end{aligned}$$

Now,

$$\begin{aligned} \frac{\partial^2 d_i}{\partial a_i \partial \Delta}(a^*, a^*) &= - (N-1) H(a^* - \Delta)^{N-2} h(a^* - \Delta)^2 + (N-1) H(a^* - \Delta)^{N-2} h(a^* - \Delta)^2 \\ &\quad + H(a^* - \Delta)^{N-1} h'(a^* - \Delta) = 0 \end{aligned}$$

when $a^* < \Delta$, because $H(b) = 0$ for $b < 0$. The implicit function theorem then gives that $\frac{\partial a^*}{\partial \Delta} = 0$ if $a^* < \Delta$. Because under Assumption 5, $a^*(0) < \Delta$, it follows that under the same assumption $a^*(\Delta) = a^*(0) < \Delta$.

Case 2: $\lambda N < 1$. In this case, the equilibrium fee a^* cannot solve the first order condition (3.12). As proven above, the solution to this foc satisfies $a^* \leq \Delta$. When $a^* \leq \Delta$, every consumer wants to visit at least one adviser. That is, $\sum_i d_i(a^*, a^*) = 1$. But when $\lambda N < 1$, this contradicts that $d_i \leq \lambda \forall i$, which as I note in the main text is a necessary condition for profit maximization.

Thus, any a^* satisfying the first order condition (3.12) cannot be an equilibrium. Because I have proven that a unique equilibrium exists, which moreover is symmetric, it must be on the boundary of advisers' constraint sets. The upper boundary $a^* = \Delta + \bar{b}$ is trivially not an equilibrium, for every adviser would face a demand equal to zero. The

⁴¹That duopolists charge a lower fee than a monopolist follows from Corollary 1 in Chen and Riordan (2008). That fees decrease further as the number of advisers increases from two follows from Theorem 2 in Quint (2014).

only remaining possibility is then that in equilibrium, every adviser operates exactly at capacity. Thus, the unique equilibrium satisfies

$$d_i(a^*, a^*) = \lambda.$$

It follows immediately that $n^* = \lambda N$.

Case 3: $\lambda N = 1$. The remaining case, $\lambda N = 1$ falls under both of the preceding cases. The first order conditions (3.12) have an admissible solution. $a^* \leq \Delta$ implies that $Nd_i(a^*, a^*) = 1$. This gives $d_i(a^*, a^*) = \lambda$, which is possible given the available capacity in the market.

Proof of Proposition 14

Assume that the market is covered, i.e. $\lambda N^* \geq 1$. As noted in the proof of Proposition 13, in this case, the equilibrium fee a^* is equal to the fee charged by advisers that deliver only convenience benefits. Since the total surplus delivered by such advisers is at most $\mathbb{E}[b_i]$, it follows that the advisers' profits is also at most $\mathbb{E}[b_i]$. It follows immediately from Assumption 4 that entry is not profitable if it leads to a situation in which $\lambda N \geq 1$. Hence, it must be the case that $\lambda N^* < 1$.

Proof of Proposition 15

The proof uses the following lemma, which establishes that the expected adviser benefit of consumers visiting an adviser increases in the number of advisers.

Lemma 4. *Let $\lambda N < 1$. Then*

$$\hat{b}_N \equiv \mathbb{E}[\max\{b_1, \dots, b_N\} | \max\{b_1, \dots, b_N\} + \Delta - a^*(N) \geq 0]$$

is increasing in N . Here, I use the notation $a^(N)$ to make explicit that the equilibrium advice depends on the number of advisers N .*

Proof. Using that the cdf of $\max\{b_1, \dots, b_N\}$ conditional on $\max\{b_1, \dots, b_N\} \geq a^* - \Delta$ is

$$\frac{H(b)^N}{1 - H(a^* - \Delta)^N}$$

it follows that

$$\hat{b}_N = \frac{N}{1 - H(a^* - \Delta)^N} \int_{a^* - \Delta}^{\bar{b}} b H(b)^{N-1} h(b) db = \frac{1}{\lambda} \int_{a^* - \Delta}^{\bar{b}} b H(b)^{N-1} h(b) db.$$

Then

$$\frac{\partial \hat{b}_N}{\partial N} = \frac{(N-1)}{\lambda} \int_{a^* - \Delta}^{\bar{b}} b H(b)^{N-2} h(b)^2 db > 0.$$

□

3. FEE-BASED VERSUS COMMISSION-BASED ADVICE

Let the market contain N advisers. An additional adviser enters if

$$\pi_i^A = \lambda a^* \geq E$$

while entry increases welfare if

$$\begin{aligned} & \lambda(q_L \mathbb{E}[\varepsilon_L | \varepsilon_L - p_L^* \geq \varepsilon_H - p_H^*] + q_H \mathbb{E}[\varepsilon_H | \varepsilon_H - p_H^* \geq \varepsilon_L - p_L^*]) \\ & + \lambda(N+1) \mathbb{E}[\max\{b_1, \dots, b_N, b_{N+1}\} | \max\{b_1, \dots, b_N, b_{N+1}\} + \Delta - a^*(N+1) \geq 0] \\ & - \lambda N \mathbb{E}[\max\{b_1, \dots, b_N\} | \max\{b_1, \dots, b_N\} + \Delta - a^*(N) \geq 0] \geq E. \end{aligned}$$

Using that $\lambda(q_L \mathbb{E}[\varepsilon_L | \varepsilon_L - p_L^* \geq \varepsilon_H - p_H^*] + q_H \mathbb{E}[\varepsilon_H | \varepsilon_H - p_H^* \geq \varepsilon_L - p_L^*]) = \Delta + q_L p_L^* + q_H p_H^*$ as well as the notation \hat{b}_N introduced in the Lemma above, this condition becomes

$$\lambda(\Delta + q_L p_L^* + q_H p_H^* + (N+1)\hat{b}_{N+1} - N\hat{b}_N) \geq E.$$

Note that

$$\lambda(\Delta + q_L p_L^* + q_H p_H^* + (N+1)\hat{b}_{N+1} - N\hat{b}_N) \geq \lambda(\Delta + \hat{b}_{N+1}) \geq \lambda a_{N+1}^*$$

The first inequality uses the Lemma derived above, while the second inequality uses the fact that, by definition, $\hat{b}_{N+1} \geq a_{N+1}^* - \Delta$. Since the first term of the inequality gives the lowest value of N for which adviser entry increases welfare and the last term the lowest value of N for which adviser entry is profitable, there is insufficient entry of advisers.

Proof of Lemma 2

The proof is very similar to that of Lemma 1. I give the main differences. First, passive beliefs imply that a consumer infers from a recommendation to purchase $i \in \{L, H\}$ that (for $j \neq i$)

$$k_i^* + \gamma(\varepsilon_i - p_i) \geq k_j^* + \gamma(\varepsilon_j - p_j)$$

as well as

$$k_i^* + \gamma(\varepsilon_i - p_i) \geq 0.$$

Therefore,

$$\begin{aligned} \mathbb{E}[\varepsilon_i - p_i | i \text{ recommended}] &= \rho_L \mathbb{E} \left[\varepsilon_i - p_i \middle| \varepsilon_i - p_i \geq \max \left\{ \varepsilon_j - p_j + \frac{k_H^* - k_L^*}{\gamma}, -\frac{k_L^*}{\gamma} \right\} \right] \\ &+ \rho_H \mathbb{E} \left[\varepsilon_i - p_i \middle| \varepsilon_i - p_i \geq \max \left\{ \varepsilon_j - p_j + \frac{k_L^* - k_H^*}{\gamma}, -\frac{k_H^*}{\gamma} \right\} \right], \end{aligned}$$

where ρ_L and ρ_H are the probabilities that the adviser recommends product L and H respectively. Similarly,

$$\begin{aligned} \mathbb{E}[\varepsilon_j - p_j | i \text{ recommended}] = & \rho_L \mathbb{E} \left[\varepsilon_j - p_j \middle| \varepsilon_i - p_i \geq \max \left\{ \varepsilon_j - p_j + \frac{k_H^* - k_L^*}{\gamma}, -\frac{k_L^*}{\gamma} \right\} \right] \\ & + \rho_H \mathbb{E} \left[\varepsilon_j - p_j \middle| \varepsilon_i - p_i \geq \max \left\{ \varepsilon_j - p_j + \frac{k_L^* - k_H^*}{\gamma}, -\frac{k_H^*}{\gamma} \right\} \right], \end{aligned}$$

Similar steps as in Lemma 1 can be used to show that

$$\mathbb{E}[\varepsilon_i - p_i | i \text{ recommended}] \geq \mathbb{E}[\varepsilon_j - p_j | i \text{ recommended}].$$

In addition, it is required that the consumer prefers purchasing i over the outside option. Therefore, the consumer follows the recommendation to purchase product i if

$$\hat{\varepsilon}_i(p_i) \equiv \mathbb{E}[\varepsilon_i - p_i | i \text{ recommended}] \geq 0.$$

When this does not hold, the consumer prefers the outside option over product i . Transitivity implies the consumer prefers the outside option over product j as well. Therefore, the consumer makes no purchase when an adviser recommends product i and $\hat{\varepsilon}_i(p_i) < 0$.

Finally, I have to prove that a consumer always follows an adviser's recommendation to purchase the outside option. When the adviser recommends the outside option, the consumer infers that

$$\max\{k_L^* + \gamma(\varepsilon_L - p_L), k_H^* + \gamma(\varepsilon_H - p_H)\} < 0.$$

Since kickbacks are positive, this immediately implies that

$$\varepsilon_i - p_i < 0$$

for $i \in \{L, H\}$ and the consumer prefers the outside option over either product.

Proof of Proposition 17

I start with existence and uniqueness. As noted in the main text, the kickback game is a special case of a game analyzed by Quint (2014). In his model, every product is allowed to consist of several complementary “parts”, which are priced by independent monopolists. If every product consists of exactly one part, his model is a model of gross substitutes. Lemma 1 in Quint (2014) proves that the equilibrium of this game exists and is unique, as well as that the game is supermodular. He assumes that (in the notation of this paper) $G(\varepsilon)$ and $1 - G(\varepsilon)$ are log-concave. As noted by Bagnoli and Bergstrom

(2005), this is weaker than the density $g(\varepsilon)$ being log-concave and differentiable. He further assumes that ε has unbounded support (Assumption 2). This assumption is only needed to rule out certain “no-trade” equilibria that can occur when the model features complementaries. Since in my particular case there are no complementaries, this assumption is not needed here.

By Theorem 1 in Quint (2014), demand is log-concave. Therefore, a firm’s profit function is quasi-concave and $k_L^* > k_H^*$ can be proven in the same way as in Proposition 6.

Proof of Proposition 18

Consider $s = 0$. First consider consumers’ choice of channel under direct remuneration. As noted in the main text, a consumer’s expected surplus when searching is

$$\mathbb{E} \left[\max \left\{ 0, \max_{i \in \{L, H\}} \{ \min\{\varepsilon_i, \varepsilon^*\} - p_i \} \right\} \right], \quad (3.13)$$

while the consumer’s expected surplus at the adviser is

$$\mathbb{E} \left[\max \left\{ 0, \max_{i \in \{L, H\}} \varepsilon_i - p_i \right\} \right].$$

Since $s = 0$ implies $\varepsilon^* = 0$, it follows immediately from comparing these quantities that every consumer weakly prefers going to an adviser to searching himself, even if the maximum convenience benefit offered by advisers equals zero and advice is free.

Under kickbacks, consumers’ expected surplus of searching is also given by (3.13). The expected consumer surplus of visiting an adviser is

$$\mathbb{E} \left[\max_{i \in \{L, H\}} \varepsilon_i - p_i \middle| k^* + \gamma \left(\max_{i \in \{L, H\}} \varepsilon_i - p_i \right) \geq 0 \right].$$

Since $k^* > 0$, this is strictly smaller than (3.13) when $\varepsilon^* = 0$ so that a consumer only visits an adviser if the convenience benefits it offers is sufficiently large. It follows, that when $s = 0$ and convenience benefits are sufficiently small, all consumers search themselves under kickbacks.

Therefore, for both remuneration schemes, consumers purchase good i if

$$P(\varepsilon_i - p_i \geq \max\{0, \varepsilon_j - p_j\}).$$

It follows that equilibrium prices under kickbacks and under direct remuneration are the same. However, because consumers going to an adviser receive some additional surplus from adviser benefits, consumers are strictly better off under direct remuneration.

The equilibrium can be written as the solution to a set of equations (firms' first order conditions and the entry conditions). Since these equations are continuous, their solutions (i.e. prices and number of entrants, ignoring the integer constraint) are also continuous by the implicit function theorem. Because consumer surplus is continuous in prices, the number of firms and \bar{b} , and because consumers are strictly better off when $s = 0$, it follows that for s and \bar{b} sufficiently small, consumers are better off when advisers are paid directly by consumers.

3.B Alternative assumptions

3.B.1 The model without capacity constraints

In this appendix, I derive the model when advisers do not face capacity constraints. Note that the short-run results in Section 3.3 are already derived under this assumption. Therefore, I only revisit the long-run results in Section 3.4. I do this under two simplifying assumptions. The first one is that the firms are symmetric, i.e. $c = 0$. The second one is that match values have uniform distributions, i.e. $\varepsilon_i \sim \text{Unif}(0, \bar{\varepsilon})$.

Under these assumptions, it is easy to calculate that for commission-based remuneration the equilibrium is

$$p^* = \frac{2}{3}\bar{\varepsilon}, \quad k^* = p^* - \frac{\gamma}{2}\bar{\varepsilon}.$$

Per-adviser profit is then $\pi_j^A = \frac{k^*}{N}$. Assume that at least one adviser wants to enter, i.e.

$$E \leq k^* = \frac{(4 - 3\gamma)\bar{\varepsilon}}{6}. \quad (3.14)$$

Then, the number of advisers is

$$N_k^* = \frac{k^*}{E} = \frac{(4 - 3\gamma)\bar{\varepsilon}}{6E}.$$

Under direct remuneration, it follows from the first order conditions that the firms set price $p^* = \frac{\bar{\varepsilon}}{2}$. Expected consumer surplus from purchasing the recommended product is then

$$\Delta = \mathbb{E}[\max\{\varepsilon_L, \varepsilon_H\}] - p^* = \frac{2}{3}\bar{\varepsilon} - \frac{1}{2}\bar{\varepsilon} = \frac{\bar{\varepsilon}}{6}.$$

The profit of adviser i is $\pi_i^A(a_i, a_{-i}) = a_i d_i(a_i, a_{-i})$, with demand

$$d_i(a_i, a_{-i}) = P \left(\underbrace{b_i + \Delta - a_i \geq 0}_{\text{preferred over no advice}}, \underbrace{b_i - a_i \geq \max_j b_j - a_j}_{\text{preferred over other advisers}} \right).$$

3. FEE-BASED VERSUS COMMISSION-BASED ADVICE

By a similar reasoning as in the proof of Proposition 13, there exists a unique equilibrium. Because the advisers are symmetric, this equilibrium is symmetric and given by the first order condition

$$\frac{\partial \pi_i^A}{\partial a_i}(a^*, a^*) = 0.$$

Using the assumption that $b_i \sim \text{Unif}(0, \bar{b})$, the demand for a deviant adviser i can be written as

$$d_i(a_i, a^*) = \int_{\max\{a_i - \Delta, a_i - a^*, 0\}}^{\bar{b}} \left(\frac{b + a^* - a_i}{\bar{b}} \right)^{N-1} \frac{1}{\bar{b}} db.$$

By a similar reasoning as for the case $\lambda N > 1$ in the proof of Proposition 13, Assumption 5 implies that $a^* < \Delta$. Therefore, the first order condition reduces to

$$-\frac{a^*}{\bar{b}} + \frac{1}{N} = 0,$$

so that the adviser fees are $a^* = \frac{\bar{b}}{N}$.

In equilibrium, the demand for every adviser is $d_i(a^*, a^*) = \frac{1}{N}$, so that $\pi_i^A = \frac{\bar{b}}{N^2}$. Therefore, under fee-based remuneration, the long-run number of advisers is

$$N_d^* = \sqrt{\frac{\bar{b}}{E}}.$$

Therefore, there are more advisers under commission-based remuneration if and only if

$$\bar{b}E \leq \left(\frac{(4 - 3\gamma)\bar{\varepsilon}}{6} \right)^2.$$

Assumption 4 implies that $\bar{b} < E$. Therefore, the left-hand side is smaller than E^2 . But under the assumption that at least one adviser wants to enter, i.e. assuming (3.14) holds, E^2 is smaller than the right-hand side. Therefore, more advisers enter under commission-based advice than under fee-based advice.

However, as long as \bar{b} is small enough, consumers are now better off without kickbacks. To see why, note that the expected consumer surplus under kickbacks is

$$CS_k = \bar{\varepsilon} - p^* + \mathbb{E}[\max_{1 \leq i \leq N_k^*} b_i] = \frac{N_k^*}{N_k^* + 1} \bar{b}.$$

When \bar{b} is small enough, $\Delta \geq a^*$ so that every consumer visits an adviser. Hence, without kickbacks expected consumer surplus is

$$CS_d = \Delta + \mathbb{E}[\max_{1 \leq i \leq N_d^*} b_i] - a^* = \frac{\bar{\varepsilon}}{6} + \frac{N_d^*}{N_d^* + 1} \bar{b} - \frac{\bar{b}}{N_d^*}.$$

It is immediate that if \bar{b} is sufficiently small, $CS_d > CS_k$.

3.B.2 Long-run equilibria without advisers' convenience benefits

In this appendix, I consider the model with homogeneous advisers. That is, advisers do not offer convenience benefits, or $\bar{b} = 0$. Without capacity constraints, the analysis remains essentially the same as in Section 3.3.

With capacity constraints, the resulting equilibrium is quite different than the one derived in the main text. The game between advisers is now one of Bertrand-Edgeworth competition, so that a pure-strategy equilibrium does not exist. Fonseca and Normann (2013) study exactly my setting and prove that for a given number of advisers, the equilibrium can be characterized as follows.

Proposition 20 (Fonseca and Normann (2013)). *There exists a unique symmetric equilibrium for the game between advisers. When $\lambda N \leq 1$, advisers charge a fee $a^* = \Delta$. When $\lambda N \geq 1$, advisers charge a fee $a^* = 0$. For $\lambda N \in (1, 1 + \lambda)$, advisers play a mixed strategy, on the support $[\underline{a}, \Delta]$, with probability function $K(a)$, where*

$$\underline{a} = \frac{1 - (N - 1)\lambda}{\lambda} \Delta,$$

$$K(a) = \left(\frac{\lambda a - (1 - (N - 1)\lambda)\Delta}{(\lambda N - 1)a} \right)^{\frac{1}{N-1}}.$$

Expected profits are as follows:

$$\mathbb{E}[\pi^A] = \begin{cases} \Delta & \text{if } \lambda N \leq 1, \\ \frac{1 - \lambda(N-1)}{\lambda} \Delta & \text{if } \lambda N \in (1, 1 + \lambda), \\ 0 & \text{if } \lambda N \geq 1 + \lambda. \end{cases}$$

When $\Delta > E$, at least one adviser enters. In fact, advisers enter as long as for the last entrant $\lambda N < 1$. From this point, at most one more adviser enters; if two more advisers enter, $\lambda N \geq 1$ and the last entrant makes no profits. In other words, it must either be the case that $N = \lfloor 1/\lambda \rfloor$ or $N = \lfloor 1/\lambda \rfloor + 1$, where $\lfloor \cdot \rfloor$ is the floor operator. Which case occurs depends on the remainder $1 \bmod \lambda$. To see this, assume that $N = \lfloor 1/\lambda \rfloor$ and consider the decision of a potential entrant. When $1 \bmod \lambda$ is sufficiently close to one, $\lambda(N + 1)$ is only slightly higher than one. The expected profits of the entrant are therefore close to Δ and this adviser wants to enter. In this case, consumers expect to pay less than Δ for advice, so that they receive positive surplus from visiting an adviser. When $1 \bmod \lambda$ is close to zero however, $\lambda(N + 1)$ will be close to $1 + \lambda$, so that the expected profits of the potential entrant are close to zero as well and it stays out of the market. In this case, consumers pay Δ for advice so that their expected surplus is zero.

Summary

In the textbook economic model, consumers always purchase the product that fits them the best. However, many markets have features that prevent consumers from obtaining their best match. This thesis uses recent changes to the Dutch mortgage market to gain a better understanding of such frictional markets. By doing so, it contributes to a wider academic literature that aims to understand the effects of such frictions and what policy makers can do about them.

Chapter 1 empirically studies the prohibition of history-based price discrimination in the Dutch mortgage market. History-based price discrimination is a strategy firms employ when it is costly for consumers to switch to a different provider. Under this type of price discrimination, a firm's existing customers are charged higher prices than new customers as the first group is "locked in" due to switching costs. It finds that for an average mortgage, banning history-based price discrimination increases welfare by €125 per year and consumer surplus by €415 per year, while bank profits drop by €290 per year.

Chapter 2 studies the effect of competition in markets with switching costs. Theoretically, I show that an increase in competition, as measured by the number of firms active in a market, amplifies the dynamic pricing incentives firms have in markets with switching costs. Empirical support for the theory is found by studying recent entry by pension funds into the Dutch mortgage market.

Chapter 3 theoretically studies the regulation of financial advisers. Some regulators have banned commission payments to financial advisers, because they might lead to biased advice. When commissions are banned, advisers charge consumers a fixed fee. To investigate when fee-based advice is preferable to commission-based advice, this chapter builds a theoretical model of advice that takes into account the entry and exit of advisers. For a fixed number of advisers, a ban on commissions increases consumer

SUMMARY

surplus because advisers are no longer biased. The ban however hurts the profitability of advisers, so that in the long run, they exit the market, advice becomes inaccessible and the ban no longer benefits consumers. These results can explain why commission bans might cause an “advice gap” and imply that accounting for the endogenous structure of the market is important when regulating advice.

Bibliography

- Abadie, Alberto, Susan Athey, Guido W Imbens, and Jeffrey Wooldridge. 2017a. "When should you adjust standard errors for clustering?" *Mimeo*.
- Abadie, Alberto, Susan Athey, Guido W Imbens, and Jeffrey M Wooldridge. 2017b. "Sampling-based vs. design-based uncertainty in regression analysis." *Mimeo*.
- Abbring, Jaap H., Jeffrey R. Campbell, Jan Tilly, and Nan Yang. 2017. "Very Simple Markov-Perfect Industry Dynamics: Empirics." *Mimeo*.
- Abe, Naohito, and Daiji Kawaguchi. 2010. "Incumbent's price response to new entry: the case of Japanese supermarkets." *Journal of the Japanese and International Economies* 24 (2): 196–212.
- Ackerberg, Daniel, C. Lanier Benkard, Steven Berry, and Ariel Pakes. 2007. "Econometric tools for analyzing market outcomes." *Handbook of Econometrics* 6:4171–4276.
- Aguirregabiria, Victor, and Pedro Mira. 2007. "Sequential estimation of dynamic discrete games." *Econometrica* 75 (1): 1–53.
- Ailawadi, Kusum L, Jie Zhang, Aradhna Krishna, and Michael W Kruger. 2010. "When Wal-Mart enters: How incumbent retailers react and how this affects their sales outcomes." *Journal of Marketing Research* 47 (4): 577–593.
- Alé, Jorge. 2013. "Switching Costs and Introductory Pricing in the Wireless Service Industry." *Mimeo*.
- Allen, Jason, Robert Clark, and Jean-François Houde. 2014. "The effect of mergers in search markets: Evidence from the Canadian mortgage industry." *American Economic Review* 104 (10): 3365–96.

BIBLIOGRAPHY

- Arie, Guy, and Paul LE Grieco. 2014. "Who pays for switching costs?" *Quantitative Marketing and Economics* 12 (4): 379–419.
- Armstrong, Mark, and Jidong Zhou. 2011. "Paying for prominence." *Economic Journal* 121 (556): 368–395.
- Asker, John, Allan Collard-Wexler, and Jan De Loecker. 2017. "Market Power, Production (Mis)Allocation and OPEC." *Mimeo*.
- Asplund, Marcus, Rickard Eriksson, and Niklas Strand. 2008. "Price discrimination in oligopoly: evidence from regional newspapers." *Journal of Industrial Economics* 56 (2): 333–346.
- Athey, Susan, and Guido W Imbens. 2018. "Design-based analysis in difference-in-differences settings with staggered adoption." *Mimeo*.
- Atkin, David, Benjamin Faber, and Marco Gonzalez-Navarro. 2018. "Retail globalization and household welfare: Evidence from Mexico." *Journal of Political Economy* 126 (1): 1–73.
- Autoriteit Financiële Markten. 2015. "Rentekorting bij hypothekeken." Accessed December 8, 2016. <https://www.afm.nl/nl-nl/professionals/nieuws/2015/mei/rentekorting-hypothekeken>.
- Bagnoli, Mark, and Ted Bergstrom. 2005. "Log-concave probability and its applications." *Economic Theory* 26 (2): 445–469.
- Bain, Joe S. 1968. *Industrial Organization*. Wiley.
- Bajari, Patrick, C. Lanier Benkard, and Jonathan Levin. 2007. "Estimating dynamic models of imperfect competition." *Econometrica* 75 (5): 1331–1370.
- Bardey, David, Denis Gromb, David Martimort, and Jérôme Pouyet. 2019. "Controlling Sellers Who Provide Advice: Regulation and Competition." *CEPR Discussion Paper* 13543.
- Barone, Guglielmo, Roberto Felici, and Marcello Pagnini. 2011. "Switching costs in local credit markets." *International Journal of Industrial Organization* 29 (6): 694–704.
- Barwick, Panle Jia, and Parag A. Pathak. 2015. "The costs of free entry: an empirical study of real estate agents in Greater Boston." *RAND Journal of Economics* 46 (1): 103–145.

- Basker, Emek. 2005. "Selling a cheaper mousetrap: Wal-Mart's effect on retail prices." *Journal of Urban Economics* 58 (2): 203–229.
- Basker, Emek, and Michael Noel. 2009. "The evolving food chain: competitive effects of Wal-Mart's entry into the supermarket industry." *Journal of Economics & Management Strategy* 18 (4): 977–1009.
- Beggs, Alan, and Paul Klemperer. 1992. "Multi-period competition with switching costs." *Econometrica* 60 (3): 651–666.
- Belloni, Alexandre, and Victor Chernozhukov. 2013. "Least squares after model selection in high-dimensional sparse models." *Bernoulli* 19 (2): 521–547.
- Bennett, Daniel, and Wesley Yin. 2019. "The Market for High-Quality Medicine: Retail Chain Entry and Drug Quality in India." *Review of Economics and Statistics* 101 (1): 76–90.
- Berry, Steven, James Levinsohn, and Ariel Pakes. 1995. "Automobile prices in market equilibrium." *Econometrica* 63 (4): 841–890.
- . 2004. "Differentiated products demand systems from a combination of micro and macro data: The new car market." *Journal of Political Economy* 112 (1): 68–105.
- Beshears, John, James J. Choi, David Laibson, and Brigitte C. Madrian. 2018. "Behavioral Household Finance." *Mimeo*.
- Biglaiser, Gary, Jacques Crémer, and Gergely Dobos. 2013. "The value of switching costs." *Journal of Economic Theory* 148 (3): 935–952.
- . 2016. "Heterogeneous switching costs." *International Journal of Industrial Organization* 47:62–87.
- Bils, Mark. 1989. "Pricing in a customer market." *Quarterly Journal of Economics* 104 (4): 699–718.
- Bolton, Patrick, Xavier Freixas, and Joel Shapiro. 2007. "Conflicts of interest, information provision, and competition in the financial services industry." *Journal of Financial Economics* 85 (2): 297–330.
- Bresnahan, Timothy F, and Peter C Reiss. 1991. "Entry and competition in concentrated markets." *Journal of Political Economy* 99 (5): 977–1009.

BIBLIOGRAPHY

- Busso, Matias, and Sebastian Galiani. 2019. "The Causal Effect of Competition on Prices and Quality: Evidence from a Field Experiment." *American Economic Journal: Applied Economics* 11 (1): 33–56.
- Cabral, Luis. 2016. "Dynamic pricing in customer markets with switching costs." *Review of Economic Dynamics* 20:43–62.
- Caplin, Andrew, and Barry Nalebuff. 1991. "Aggregation and imperfect competition: On the existence of equilibrium." *Econometrica* 59 (1): 25–59.
- Carbo-Valverde, Santiago, Timothy H Hannan, and Francisco Rodriguez-Fernandez. 2011. "Exploiting old customers and attracting new ones: The case of bank deposit pricing." *European Economic Review* 55 (7): 903–915.
- Chen, Yongmin. 1997. "Paying customers to switch." *Journal of Economics & Management Strategy* 6 (4): 877–897.
- Chen, Yongmin, and Michael H Riordan. 2007. "Price and variety in the spokes model." *Economic Journal* 117 (522): 897–921.
- . 2008. "Price-increasing competition." *RAND Journal of Economics* 39 (4): 1042–1058.
- Choi, Michael, Anovia Yifan Dai, and Kyungmin Kim. 2018. "Consumer search and price competition." *Econometrica* 86 (4): 1257–1281.
- Collard-Wexler, Allan. 2013. "Demand Fluctuations in the Ready-Mix Concrete Industry." *Econometrica* 81 (3): 1003–1037.
- Corts, Kenneth S. 1998. "Third-degree price discrimination in oligopoly: all-out competition and strategic commitment." *RAND Journal of Economics* 29 (2): 306–323.
- Cosguner, Koray, Tat Y Chan, and PB Seetharaman. 2016. "Dynamic pricing in a distribution channel in the presence of switching costs." *Management Science* 64 (3): 1212–1229.
- . 2017. "Behavioral Price Discrimination in the Presence of Switching Costs." *Marketing Science* 36 (3): 426–435.
- Courtemanche, Charles, and Art Carden. 2014. "Competing with Costco and Sam's Club: Warehouse club entry and grocery prices." *Southern Economic Journal* 80 (3): 565–585.

- Cullen, Joseph, Nicolas Schutz, and Oleksandr Shcherbakov. 2015. "Welfare analysis of equilibria with and without early termination fees in the US wireless industry." *Mimeo*.
- Cullen, Joseph, and Oleksandr Shcherbakov. 2010. "Measuring consumer switching costs in the wireless industry." *Mimeo*.
- De Cornière, Alexandre, and Greg Taylor. 2016. "A Model of Biased Intermediation." *CEPR Discussion Paper 11457*.
- De Nederlandsche Bank. 2016. *Loan markets in motion: Larger role of pension funds and insurers boosts financial stability*. Technical report. Accessed May 9, 2019.
- Dijkstra, Mark A, Fleur Randag, and Maarten Pieter Schinkel. 2014a. "High mortgage rates in the low countries: an introduction." *Journal of Competition Law and Economics* 10 (4): 819–841.
- . 2014b. "High mortgage rates in the Low Countries: what happened in the Spring of 2009?" *Journal of Competition Law and Economics* 10 (4): 843–859.
- Doganoglu, Toker. 2010. "Switching costs, experience goods and dynamic price competition." *Quantitative Marketing and Economics* 8 (2): 167–205.
- Dubé, Jean-Pierre, Günter J. Hitsch, and Peter E. Rossi. 2009. "Do switching costs make markets less competitive?" *Journal of Marketing research* 46 (4): 435–445.
- Edelman, Benjamin, and Julian Wright. 2015a. "Markets with Price Coherence." *HBS Working Paper 15-01*.
- . 2015b. "Price Coherence and Excessive Intermediation." *Quarterly Journal of Economics* 130 (3): 1283–1328.
- Egan, Mark. Forthcoming. "Brokers vs. Retail Investors: Conflicting Interests and Dominated Products." *Journal of Finance*.
- Ericson, Richard, and Ariel Pakes. 1995. "Markov-perfect industry dynamics: A framework for empirical work." *The Review of Economic Studies* 62 (1): 53–82.
- Fabra, Natalia, and Alfredo García. 2015. "Market structure and the competitive effects of switching costs." *Economics Letters* 126:150–155.

BIBLIOGRAPHY

- Farrell, Joseph, and Paul Klemperer. 2007. "Coordination and lock-in: Competition with switching costs and network effects." *Handbook of industrial organization* 3:1967–2072.
- Farrell, Joseph, and Carl Shapiro. 1988. "Dynamic competition with switching costs." *RAND Journal of Economics* 19 (1): 123–137.
- Fershtman, Chaim, and Ariel Pakes. 2012. "Dynamic games with asymmetric information: A framework for empirical work." *The Quarterly Journal of Economics* 127 (4): 1611–1661.
- Financial Conduct Authority. 2016. *Financial Advice Market Review: Final Report*. Technical report. Accessed October 12, 2018. <https://www.fca.org.uk/publication/corporate/famr-final-report.pdf>.
- . 2017. *Financial Advice and Guidance: Quantitative research to inform the Financial Advice Market Review (FAMR) Baseline*. Technical report. Accessed October 12, 2018. <https://www.fca.org.uk/publication/research/famr-quantitative-research.pdf>.
- Fleitas, Sebastian. 2017. "Dynamic Competition and Price Regulation When Consumers Have Inertia: Evidence from Medicare Part D." *Mimeo*.
- Fonseca, Miguel A, and Hans-Theo Normann. 2013. "Excess capacity and pricing in Bertrand–Edgeworth markets: experimental evidence." *Journal of Institutional and Theoretical Economics* 169 (2): 199–228.
- Fransman, Robin. 2017. "Nationale Hypotheek Garantie, een terugblik." *Economisch Statistische Berichten* 102 (4749): 38–42.
- Gabaix, Xavier. 2014. "A sparsity-based model of bounded rationality." *The Quarterly Journal of Economics* 129 (4): 1661–1710.
- . 2017. "Behavioral Macroeconomics Via Sparse Dynamic Programming." *Mimeo*.
- Gehrig, Thomas, Oz Shy, and Rune Stenbacka. 2012. "A welfare evaluation of history-based price discrimination." *Journal of Industry, Competition and Trade* 12 (4): 373–393.
- Gilchrist, Simon, Raphael Schoenle, Jae Sim, and Egon Zakrajšek. 2017. "Inflation dynamics during the financial crisis." *American Economic Review* 107 (3): 785–823.

- Guiso, Luigi, Andrea Pozzi, Anton Tsoy, Leonardo Gambacorta, and Paolo Mistrulli. 2018. "The Cost of Steering in Financial Markets: Evidence from the Mortgage Market." *Mimeo*.
- Handel, Benjamin R. 2013. "Adverse selection and inertia in health insurance markets: When nudging hurts." *American Economic Review* 103 (7): 2643–82.
- Hausman, Jerry, and Ephraim Leibtag. 2007. "Consumer benefits from increased competition in shopping outlets: Measuring the effect of Wal-Mart." *Journal of Applied Econometrics* 22 (7): 1157–1177.
- Ho, Chun-Yu. 2015. "Switching cost and deposit demand in china." *International Economic Review* 56 (3): 723–749.
- Holmes, Thomas J. 1989. "The effects of third-degree price discrimination in oligopoly." *The American Economic Review* 79 (1): 244–250.
- Honka, Elisabeth. 2014. "Quantifying search and switching costs in the US auto insurance industry." *RAND Journal of Economics* 45 (4): 847–884.
- Hortacsu, Ali, and Steven L. Puller. 2008. "Understanding strategic bidding in multi-unit auctions: a case study of the Texas electricity spot market." *The RAND Journal of Economics* 39 (1): 86–114.
- Hosken, Daniel S, Luke M Olson, and Loren K Smith. 2018. "Do retail mergers affect competition? Evidence from grocery retailing." *Journal of Economics & Management Strategy* 27 (1): 3–22.
- Hotelling, Harold. 1929. "Stability in Competition." *Economic Journal* 39 (153): 41–57.
- Ifrach, Bar, and Gabriel Y. Weintraub. 2017. "A Framework for Dynamic Oligopoly in Concentrated Industries." *Review of Economic Studies* 84 (3): 1106–1150.
- Inderst, Roman, and Marco Ottaviani. 2012a. "Competition through commissions and kickbacks." *American Economic Review* 102 (2): 780–809.
- . 2012b. "How (not) to pay for advice: A framework for consumer financial protection." *Journal of Financial Economics* 105, no. 2 (August): 393–411.
- Ioannidou, Vasso, and Steven Ongena. 2010. ""Time for a change": loan conditions and bank behavior when firms switch banks." *The Journal of Finance* 65 (5): 1847–1877.

BIBLIOGRAPHY

- Janssen, Aljoscha. 2018. "Switching Costs, Quality Misconceptions and Behavioral Pricing in the Pharmaceutical Market." *Mimeo*.
- Janssen, Maarten CW, and José Luis Moraga-González. 2004. "Strategic pricing, consumer search and the number of firms." *Review of Economic Studies* 71 (4): 1089–1118.
- Jensen, Martin K. 2007. "Monotone comparative statics in ordered vector spaces." *BE Journal of Theoretical Economics* 7 (1).
- Jia, Panle. 2008. "What happens when Wal-Mart comes to town: An empirical analysis of the discount retailing industry." *Econometrica* 76 (6): 1263–1316.
- Kalouptsi, Myrto. 2014. "Time to build and fluctuations in bulk shipping." *American Economic Review* 104 (2): 564–608.
- Karaca-Mandic, Pinar, and Kenneth Train. 2003. "Standard error correction in two-stage estimation with nested samples." *Econometrics Journal* 6 (2): 401–407.
- Kim, Yeorim, and Mauro Mastrogiamomo. 2019. "The effect of the Dutch Financial Assessment Framework on the mortgage investments of Pension funds and their borrowers." *Mimeo*.
- Kiss, András. 2017. "Salience and Switching." *Mimeo*.
- Klemperer, Paul. 1987a. "Entry deterrence in markets with consumer switching costs." *Economic Journal* 97:99–117.
- . 1987b. "Markets with consumer switching costs." *Quarterly Journal of Economics* 102 (2): 375–394.
- . 1987c. "The competitiveness of markets with switching costs." *RAND Journal of Economics* 18 (1): 138–150.
- . 1988. "Welfare effects of entry into markets with switching costs." *Journal of Industrial Economics* 37 (2): 159–165.
- . 1989. "Price wars caused by switching costs." *Review of Economic Studies* 56 (3): 405–420.
- Kreps, David M, and Jose A Scheinkman. 1983. "Quantity precommitment and Bertrand competition yield Cournot outcomes." *The Bell Journal of Economics*: 326–337.

- Lach, Saul, and José L Moraga-González. 2017. "Asymmetric price effects of competition." *Journal of Industrial Economics* 65 (4): 767–803.
- Lira, Loreto, Rosario Rivero, and Rodrigo Vergara. 2007. "Entry and prices: Evidence from the supermarket sector." *Review of Industrial Organization* 31 (4): 237–260.
- Liu, Jun, Shuiwang Ji, and Jieping Ye. 2009. "Multi-task feature learning via efficient l_2 , l_1 -norm minimization." In *Proceedings of the twenty-fifth conference on uncertainty in artificial intelligence*, 339–348. AUAI Press.
- MacKay, Alexander, and Marc Remer. 2018. "An Empirical Model of Consumer Affiliation and Dynamic Price Competition." *Mimeo*.
- Manski, Charles F., and Steven R. Lerman. 1977. "The estimation of choice probabilities from choice based samples." *Econometrica* 45 (8): 1977–1988.
- Manski, Charles F, and John V Pepper. 2018. "How do right-to-carry laws affect crime rates? Coping with ambiguity using bounded-variation assumptions." *Review of Economics and Statistics* 100 (2): 232–244.
- Maskin, Eric, and Jean Tirole. 2001. "Markov perfect equilibrium: I. Observable actions." *Journal of Economic Theory* 100 (2): 191–219.
- Mastrogiacomo, Mauro, and Remco Van der Molen. 2015. "Dutch mortgages in the DNB loan level data." *DNB Occasional Studies* 13 (4).
- Milgrom, Paul, and John Roberts. 1990. "Rationalizability, learning, and equilibrium in games with strategic complementarities." *Econometrica*: 1255–1277.
- Miller, Daniel P., and Jungwon Yeo. 2012. "Estimating Dynamic Discrete Choice Models of Product Differentiation: An Application to Medicare Part D with Switching Costs." *Mimeo*.
- Murooka, Takeshi. 2015. "Deception under Competitive Intermediation." *Mimeo*.
- Myśliwski, Mateusz, and May Rostom. 2018. "Value of Information and the Impact of Mortgage Intermediaries on Lender Competition and Households' Financial Positions." *Mimeo*.
- Nederlandse Mededingingsautoriteit. 2011. *Sectorstudie Hypotheekmarkt: Een onderzoek naar de concurrentieomstandigheden op de Nederlandse hypotheekmarkt*. Accessed December 8, 2016. <https://www.acm.nl/nl/publicaties/publicatie/7091/Sectorstudie-Hypotheekmarkt/>.

BIBLIOGRAPHY

- Nosal, Kathleen. 2012. "Estimating Switching Costs for Medicare Advantage Plans." *Mimeo*.
- Obozinski, Guillaume, Ben Taskar, and Michael I. Jordan. 2010. "Joint covariate selection and joint subspace selection for multiple classification problems." *Statistics and Computing* 20 (2): 231–252.
- Padilla, A Jorge. 1995. "Revisiting dynamic duopoly with consumer switching costs." *Journal of Economic Theory* 67 (2): 520–530.
- Pakes, Ariel, Michael Ostrovsky, and Steven Berry. 2007. "Simple estimators for the parameters of discrete dynamic games (with entry / exit examples)." *RAND Journal of Economics* 38 (2): 373–399.
- Pearcy, Jason. 2016. "Bargains followed by bargains: when switching costs make markets more competitive." *Journal of Economics & Management Strategy* 25 (4): 826–851.
- Perloff, Jeffrey M, and Steven C Salop. 1985. "Equilibrium with product differentiation." *Review of Economic Studies* 52 (1): 107–120.
- Pesendorfer, Martin, and Philipp Schmidt-Dengler. 2008. "Asymptotic least squares estimators for dynamic games." *Review of Economic Studies* 75 (3): 901–928.
- Petrin, Amil, and Kenneth Train. 2010. "A control function approach to endogeneity in consumer choice models." *Journal of Marketing Research* 47 (1): 3–13.
- Phelps, Edmund S., and Sydney G. Winter Jr. 1970. "Optimal Price Policy under Atomistic Price Competition." In *Microeconomic Foundations of Employment and Inflation Theory*. New York: WW Norton Co, edited by Edmund S. Phelps, 309–337. New York: W. W. Norton & Co.
- Quint, Daniel. 2014. "Imperfect competition with complements and substitutes." *Journal of Economic Theory* 152:266–290.
- Raval, Devesh, and Ted Rosenbaum. Forthcoming. "Why Do Previous Choices Matter for Hospital Demand? Decomposing Switching Costs from Unobserved Preferences." *Review of Economics and Statistics*.
- Rhodes, Andrew. 2012. "History-based price discrimination in markets with switching costs: who gains and who loses?" *Mimeo*.
- . 2014. "Re-examining the effects of switching costs." *Economic Theory* 57 (1): 161–194.

- Rickert, Dennis. 2016. "Consumer Switching Costs, Dynamic Pricing, and Antitrust Analysis. Evidence from the Market for Diapers."
- Robinson, Joan. 1933. *Economics of imperfect competition*. McMillan.
- Robles-Garcia, Claudia. 2019. "Competition and Incentives in Mortgage Markets: The Role of Brokers." *Mimeo*.
- Romano, Joseph P., Azeem M. Shaikh, and Michael Wolf. 2014. "A practical two-step method for testing moment inequalities." *Econometrica* 82 (5): 1979–2002.
- Rosenthal, Robert W. 1980. "A model in which an increase in the number of sellers leads to a higher price." *Econometrica* 48 (6): 1575–1579.
- Ruiz-Aliseda, Francisco. 2016. "When do switching costs make markets more or less competitive?" *International Journal of Industrial Organization* 47:121–151.
- Shcherbakov, Oleksandr. 2016. "Measuring consumer switching costs in the television industry." *RAND Journal of Economics* 47 (2): 366–393.
- Shen, Bo, and Julian Wright. Forthcoming. "Why (don't) firms free ride on an intermediary's advice?" *International Journal of Industrial Organization*.
- Somaini, Paulo, and Liran Einav. 2013. "A model of market power in customer markets." *Journal of Industrial Economics* 61 (4): 938–986.
- Stango, Victor. 2002. "Pricing with consumer switching costs: Evidence from the credit card market." *Journal of Industrial Economics* 50 (4): 475–492.
- Stole, Lars A. 2007. "Price Discrimination and Competition." In *Handbook of Industrial Organization*, edited by Mark Armstrong and Robert K. Porter, 2221–2299. Elsevier.
- Stoughton, Neal M, Youchang Wu, and Josef Zechner. 2011. "Intermediated Investment Management." *Journal of Finance* 66 (3): 947–980.
- Taylor, Curtis R. 2003. "Supplier surfing: Competition and consumer behavior in subscription markets." *RAND Journal of Economics*: 223–246.
- Teh, Tat-How, and Julian Wright. 2019. "Steering by information intermediaries." *Mimeo*.
- Thiel, Jurre H. 2018a. "Fee-based versus commission-based advice." *Mimeo*.
- . 2018b. "Price discrimination, switching costs and welfare: Evidence from the Dutch mortgage market." *Mimeo*.

BIBLIOGRAPHY

- Tibshirani, Robert. 1996. "Regression shrinkage and selection via the lasso." *Journal of the Royal Statistical Society. Series B (Methodological)*: 267–288.
- To, Theodore. 1996. "Multi-period competition with switching costs: An overlapping generations formulation." *Journal of Industrial Economics* 44 (1): 81–87.
- Viard, V. Brian. 2007. "Do switching costs make markets more or less competitive? The case of 800-number portability." *RAND Journal of Economics* 38 (1): 146–163.
- Weiergräber, Stefan. 2017. "Network effects and switching costs in the US wireless industry: Disentangling sources of consumer inertia." *Mimeo*.
- Weintraub, Gabriel Y., C. Lanier Benkard, and Benjamin Van Roy. 2008. "Markov perfect industry dynamics with many firms." *Econometrica* 76 (6): 1375–1411.
- Weitzman, Martin L. 1979. "Optimal Search for the Best Alternative." *Econometrica* 47 (3): 641.
- Weyl, E Glen, and Michal Fabinger. 2013. "Pass-through as an economic tool: Principles of incidence under imperfect competition." *Journal of Political Economy* 121 (3): 528–583.
- Wollmann, Thomas G. 2018. "Trucks without bailouts: Equilibrium product characteristics for commercial vehicles." *American Economic Review* 108 (6): 1364–1406.
- Zhang, Jian. 2006. "A Probabilistic Framework for Multi-Task Learning." PhD diss., Carnegie Mellon University.

The Tinbergen Institute is the Institute for Economic Research, which was founded in 1987 by the Faculties of Economics and Econometrics of the Erasmus University Rotterdam, University of Amsterdam and VU University Amsterdam. The Institute is named after the late Professor Jan Tinbergen, Dutch Nobel Prize laureate in economics in 1969. The Tinbergen Institute is located in Amsterdam and Rotterdam. The following books recently appeared in the Tinbergen Institute Research Series:

- 706 J.D. ZHENG, *Social Identity and Social Preferences: An Empirical Exploration*
- 707 B.A. LOERAKKER, *On the Role of Bonding, Emotional Leadership, and Partner Choice in Games of Cooperation and Conflict*
- 708 L. ZIEGLER, *Social Networks, Marital Sorting and Job Matching. Three Essays in Labor Economics*
- 709 M.O. HOYER, *Social Preferences and Emotions in Repeated Interactions*
- 710 N. GHEBRIHIWET, *Multinational Firms, Technology Transfer, and FDI Policy*
- 711 H.FANG, *Multivariate Density Forecast Evaluation and Nonparametric Granger Causality Testing*
- 712 Y. KANTOR, *Urban Form and the Labor Market*
- 713 R.M. TEULINGS, *Untangling Gravity*
- 714 K.J.VAN WILGENBURG, *Beliefs, Preferences and Health Insurance Behavior*
- 715 L. SWART, *Less Now or More Later? Essays on the Measurement of Time Preferences in Economic Experiments*
- 716 D. NIBBERING, *The Gains from Dimensionality*
- 717 V. HOORNWEG, *A Tradeoff in Econometrics*
- 718 S. KUCINSKAS, *Essays in Financial Economics*
- 719 O. FURTUNA, *Fiscal Austerity and Risk Sharing in Advanced Economies*
- 720 E. JAKUCIONYTE, *The Macroeconomic Consequences of Carry Trade Gone Wrong and Borrower Protection*
- 721 M. LI, *Essays on Time Series Models with Unobserved Components and Their Applications*
- 722 N. CIURILĂ, *Risk Sharing Properties and Labor Supply Disincentives of Pay-As-You-Go Pension Systems*
- 723 N.M. BOSCH, *Empirical Studies on Tax Incentives and Labour Market Behaviour*
- 724 S.D. JAGAU, *Listen to the Sirens: Understanding Psychological Mechanisms with Theory and Experimental Tests*
- 725 S. ALBRECHT, *Empirical Studies in Labour and Migration Economics*
- 726 Y.ZHU, *On the Effects of CEO Compensation*
- 727 S. XIA, *Essays on Markets for CEOs and Financial Analysts*

- 728 I. SAKALAUŠKAITE, *Essays on Malpractice in Finance*
- 729 M.M. GARDBERG, *Financial Integration and Global Imbalances*
- 730 U. THÜMMEL, *Of Machines and Men: Optimal Redistributive Policies under Technological Change*
- 731 B.J.L. KEIJSERS, *Essays in Applied Time Series Analysis*
- 732 G. CIMINELLI, *Essays on Macroeconomic Policies after the Crisis*
- 733 Z.M. LI, *Econometric Analysis of High-frequency Market Microstructure*
- 734 C.M. OOSTERVEEN, *Education Design Matters*
- 735 S.C. BARENDSE, *In and Outside the Tails: Making and Evaluating Forecasts*
- 736 S. SÓVÁGÓ, *Where to Go Next? Essays on the Economics of School Choice*
- 737 M. HENNEQUIN, *Expectations and Bubbles in Asset Market Experiments*
- 738 M.W. ADLER, *The Economics of Roads: Congestion, Public Transit and Accident Management*
- 739 R.J. DÖTTLING, *Essays in Financial Economics*
- 740 E.S. ZWIERS, *About Family and Fate: Childhood Circumstances and Human Capital Formation*
- 741 Y.M. KUTLUAY, *The Value of (Avoiding) Malaria*
- 742 A. BOROWSKA, *Methods for Accurate and Efficient Bayesian Analysis of Time Series*
- 743 B. HU, *The Amazon Business Model, the Platform Economy and Executive Compensation: Three Essays in Search Theory*
- 744 R.C. SPERNA WEILAND, *Essays on Macro-Financial Risks*
- 745 P.M. GOLEC, *Essays in Financial Economics*
- 746 M.N. SOUVERIJN, *Incentives at work*
- 747 M.H. COVENEY, *Modern Imperatives: Essays on Education and Health Policy*
- 748 P. VAN BRUGGEN, *On Measuring Preferences*
- 749 M.H.C. NIENTKER, *On the Stability of Stochastic Dynamic Systems and their use in Econometrics*
- 750 S. GARCIA MANDICÓ, *Social Insurance, Labor Supply and Intra-Household Spillovers*
- 751 Y. SUN, *Consumer Search and Quality*
- 752 I. KERKEMEZOS, *On the Dynamics of (Anti) Competitive Behaviour in the Airline Industry*
- 753 G.W. GOY, *Modern Challenges to Monetary Policy*
- 754 A.C. VAN VLODROP, *Essays on Modeling Time-Varying Parameters*
- 755 J. SUN, *Tell Me How To Vote, Understanding the Role of Media in Modern Elections*